

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:37 ; Search time 51.33 Seconds  
(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-5  
Perfect score: 88  
Sequence: 1 CECNIRKVDVNDNFP 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36:\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	100.0	999	1 R30742	Human pemphigus vu
2	88	100.0	15	1 W04845	Self epitope of de
3	83	94.3	614	1 W07908	Pemphigus vulgaris
4	74	84.1	778	1 W15489	Pemphigus foliaceu
5	60	68.2	787	1 R86865	Human protocadheri
6	52	59.1	832	1 W74089	Human HPT-1 protei
7	50	56.8	43	1 R58865	Rat-224 cadherin p
8	50	56.8	43	1 R87107	Protocadherin clon
9	50	56.8	1180	1 W61273	Homo sapiens proto
10	48	54.5	41	1 R58875	Rat-352 cadherin p
11	48	54.5	41	1 R87117	Protocadherin clon
12	48	54.5	780	1 W13130	Full length human
13	48	54.5	780	1 W25634	Human cadherin-5.
14	46	52.3	799	1 W13126	Full length rat ca
15	46	52.3	793	1 W13132	Full length human
16	46	52.3	532	1 W13127	Truncated rat cadh
17	46	52.3	793	1 W25635	Human cadherin-8.
18	46	52.3	799	1 W25640	Rat full length ca
19	46	52.3	532	1 W25640	Rat truncated cadh
20	46	52.3	560	1 W13009	Segment of desmoso
21	45	51.1	836	1 R58912	Product of alterna
22	45	51.1	682	1 R49144	Human protocadheri
23	45	51.1	904	1 R58907	Human protocadheri
24	45	51.1	43	1 R58882	Mouse-321 cadherin
25	45	51.1	43	1 R58886	Human-11 cadherin
26	45	51.1	148	1 R58876	Rat-511 cadherin p
27	45	51.1	40	1 R58878	Rat-511 cadherin p
28	45	51.1	43	1 R58868	Rat-314 cadherin p
29	45	51.1	682	1 R87154	Alternatively spli
30	45	51.1	836	1 R87153	Alternatively spli
31	45	51.1	904	1 R87147	Protocadherin clon
32	45	51.1	43	1 R87121	Protocadherin clon
33	45	51.1	41	1 R87120	Protocadherin clon
34	45	51.1	148	1 R87118	Protocadherin clon
35	45	51.1	43	1 R87125	Protocadherin clon
36	45	51.1	43	1 R87110	Protocadherin clon
37	45	51.1	794	1 W13135	Putative human cad
38	45	51.1	794	1 W25637	Human cadherin-12.
39	45	51.1	896	1 W41252	Xenopus paraxial p

40	44	50.0	391	1 R43564	Human cadherin-6.
41	44	50.0	1203	1 R58911	Product of alterna
42	44	50.0	1026	1 R58906	Human protocadheri
43	44	50.0	1203	1 R87152	Alternatively spli
44	44	50.0	797	1 R86866	Human protocadheri
45	44	50.0	1026	1 R87146	Protocadherin clon

## ALIGNMENTS

RESULT 1  
ID R30742  
AC R30742; standard; Protein: 999 AA.  
DE 14-JUN-1993 (first entry)  
DE Human pemphigus vulgaris 130KD antigen.  
KW pemphigus vulgaris; skin disease; autoantibodies;  
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.  
OS Homo sapiens.  
PN US7798918-A.  
PD 15-DEC-1992.  
PE 27-NOV-1991; 798918.  
PR 27-NOV-1991; US-798918.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
PI Amagai M, Klaus-Kovtun V, Stanley JR.  
DR N-PSDB; Q35992.  
DR N-PSDB; Q35992.  
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for  
PT diagnostic and therapeutic uses  
PS Disclosure, Fig 7, 50pp; English.  
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein  
CC and its encoding DNA may be used in the diagnosis and treatment of  
CC pemphigus vulgaris. It is thought that the antigen may be a cell  
CC adhesion molecule.  
SQ Sequence 999 AA;

Query Match 100.0%; Score 88; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1.7e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CECNIRKVDVNDNFP 15  
DB 251 CECNIRKVDVNDNFP 265

RESULT 2  
ID W04845  
AC W04845; standard; peptide; 15 AA.  
DE 18-FEB-1997 (first entry)  
DE Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerisation; self-epitope; antigen; autoimmune disease;  
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphonamidonuclease;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW influenza; haemagglutinin; reovirus; sigma protein.  
OS Homo sapiens.  
PN M09627387-A1.  
PD 12-SEP-1996.  
PE 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PA (HARD) HARVARD COLLEGE.  
PI Strominger JL, Mucherpfennig KW;  
DR WPI; 96-425218/42.  
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for  
PT identification of other self and non-self antigens implicated in  
PT auto-immune disease  
PS Claim 1; Page 40; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise

CC either an isolated human non-collagen or non-myosin basic protein  
 CC (MBP) polypeptide which is capable of tolerising an individual to an  
 CC autoantigen; or an isolated human pathogen polypeptide capable of  
 CC tolerising an individual to that polypeptide. In both cases, the  
 CC polypeptide (whether self or non-self) includes an amino acid  
 CC sequence corresponding to a sequence motif for a MHC class II  
 CC protein, such as HLA-DR, which is associated with a human autoimmune  
 CC disease and which binds to the polypeptide to activate autoreactive  
 CC T-cells in individuals with the autoimmune disease. This peptide is  
 CC derived from the human desmoglein 3 protein (amino acids 251-265)  
 CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
 CC derived from the human desmoglein protein are described in W04841-47.  
 SO Sequence 15 AA.

Query Match 100.0%; Score 88; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CECNIKVDVNDNFP 15  
 DB 1 CECNIKVDVNDNFP 15

## RESULT 3

ID W07908 standard; protein; 614 AA.

AC W07908;  
 DT 29-JAN-1997 (first entry)  
 DE Pemphigus vulgaris antigen protein extracellular region.  
 KW Autoantibody; immunoglobulin G; IgG; fusion protein; diagnosis;  
 KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;  
 KW dermatology.  
 OS Homo sapiens.  
 PN J08188540-A.  
 PD 23-JUL-1996.  
 PF 30-JUN-1995; 165632.  
 PR 30-JUN-1994; JP-173291.  
 RA (NISH/) NISHIKAWA T.  
 DR WPI; 96-388562/39.  
 PT Fused protein recognised by pemphigus vulgaris auto:antibody -  
 PT useful to treat and diagnose pemphigus vulgaris  
 PS Claim 1: Page 7-9; 9pp; Japanese  
 CC W07908 represents the human pemphigus vulgaris (PV) antigen  
 CC extracellular region. The PV antigen is produced in patients with  
 CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare  
 CC relapsing disease causing suprabasal, intra-epidermal bullae  
 CC (vesicles) of the skin and mucous membranes, which is fatal if  
 CC untreated. The PV antigen was fused to a human IgG1 hinge region  
 CC and the resulting fusion protein is useful to treat or diagnose  
 CC pemphigus vulgaris.  
 SO Sequence 614 AA;

Query Match 94.3%; Score 83; DB 1; Length 614;  
 Best Local Similarity 93.3%; Pred. No. 6.4e-06;  
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 CECNIKVDVNDNFP 15  
 DB 250 CECNIKVDVNDNFP 264

## RESULT 4

ID W15489 standard; Protein; 778 AA.

AC W15489;  
 DT 17-JUN-1997 (first entry)  
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.  
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
 KW extracellular region; antigen; hinge portion; skin;  
 KW dermatitis herpetiformis; fusion protein; detection; ss.  
 OS Chimeric - Homo sapiens.

FT Key Location/Qualifiers  
 FT domain 1. 345  
 PN J09077800-A.  
 PD 25-MAR-1997.  
 PF 12-SEP-1995; 260899.  
 PR 12-SEP-1995; JP-260899.  
 RA (NISH/) NISHIKAWA T.  
 DR WPI; 97-241758/22.  
 DR P-SDS; 766428.

PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked  
 PT through the hinge region used to treat pemphigus foliaceus  
 PS Claim 1; Page 10-12; 17pp; Japanese.  
 CC This sequence represents a fused protein recognised by pemphigus  
 CC foliaceus patient autoantibody which comprises the constant region  
 CC of IgG linked to the extracellular region of pemphigus foliaceus  
 CC antigen protein through the hinge portion. Pemphigus foliaceus is  
 CC a chronic, generalised, vesicular and scaling skin eruption similar  
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion  
 CC protein is useful to treat pemphigus foliaceus. The antigen is  
 CC especially administered through an adsorbent upon which the fusion  
 CC protein is immobilised via a carrier. The fusion protein is also  
 CC useful for detecting pemphigus foliaceus antibodies which is useful  
 CC in immunodiagnosis. The fusion protein has little or no side effects.  
 SO Sequence 778 AA;

Query Match 84.1%; Score 74; DB 1; Length 778;  
 Best Local Similarity 80.0%; Pred. No. 0.00025;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CECNIKVDVNDNFP 15  
 DB 253 CECNIKVDVNDNFP 267

## RESULT 5

ID R68665 standard; Protein; 787 AA.

AC R68665;  
 DT 27-AUG-1996 (first entry)  
 DE Human protocadherin pc3.  
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;  
 KW catenin; therapy.  
 OS Homo sapiens.  
 PN W09600289-A1.  
 PD 04-JAN-1996.  
 PF 26-JUN-1995; U08071.  
 PR 27-JUN-1994; US-268161.  
 RA (DOHE-) DOHERTY EYE INST.  
 DR Suzuki S.  
 DR WPI; 96-068873/07.  
 DR N-PSDB; T03572.

PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat  
 PT pc5 - involved in cell-cell adhesion and regulation activities  
 PS Claim 15; Page 115-119; 146pp; English.  
 CC R68665-R68667 represents the human protocadherin pc3. These sequences are  
 CC related to cadherin, and possess cell adhesive ability. Cadherins are  
 CC glycosylated integral membrane proteins that are involved in cell-cell  
 CC adhesion. Cadherins are composed of an N-terminal extracellular domain  
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a  
 CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the  
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.  
 CC The cytoplasmic domain is not present in all cadherins, but in those  
 CC which possess it, it is essential for the cadherins adhesive function.  
 CC The cadherins which do not possess a cytoplasmic domain appear to  
 CC function via a different method from those with a cytoplasmic domain.  
 CC These sequences were isolated using primers 1 and 2 (see T03575 and  
 CC T03576). The proteins may have regulatory functions in the cell, as well  
 CC as the cell-cell adhesive properties. Antibodies produced against these  
 CC sequences are useful for modulating the binding activity of these  
 CC protocadherins, and can be used therapeutically.

SQ Sequence 787 AA;  
 Query Match 68.2%; Score 60; DB 1; Length 787;  
 Best Local Similarity 66.7%; Pred. No. 0.049;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 CECNIRKVDVNDNFP 15  
 DB 331 CSVSXKVLVDVNDNFP 345  
 RESULT 6  
 W74089  
 ID W74089 standard; Protein: 832 AA.  
 AC W74089;  
 DT 04-MAY-1999 (first entry)  
 DE Human HPT-1 protein sequence.  
 KW Gastro-intestinal transport receptor; binding protein; hsi; HPT1;  
 KM D2H: hPEPT1; human; GI tract receptor; sucrose-isomaltase complex;  
 KM intestinal peptide associated transporter; hypertension; diabetes;  
 KM osteoporosis; haemophilia; anaemia; cancer; migraine; angina pectoris;  
 KM therapeutic agent delivery; therapy; probe.  
 OS Homo sapiens.  
 PN MO9851325-A2.  
 PD 19-NOV-1998.  
 PR 15-MAY-1998; U10088.  
 PR 15-MAY-1997; US-046595.  
 PA (CYTO-) CYTOGEN CORP.  
 PA (ELAN-) ELAN CORP. PLC.  
 PI Alvarez VL, Belinka BA, Cagney GW, Carter JW, Lambkin LJ,  
 PI Omahony DJ, Patterson CA, Singleton J;  
 DR WPI: 99-009568/01.  
 DR N-PSDB: X18166.  
 PT New proteins that bind specifically to receptors in the  
 PT gastro-intestinal tract and related nucleic acid - chimeras and  
 PT antibodies, used to deliver therapeutic or diagnostic agents to, or  
 PT through, the gastrointestinal tract, e.g. insulin or leuprolide  
 PS Disclosure: fig 2; 294pp; English.  
 CC This sequence is the human HPT-1 protein. The invention relates to  
 CC purified proteins (I) that bind specifically to at least one of the  
 CC gastro-intestinal (GI) tract receptors human intestinal  
 CC peptide-associated transporter (HPT1), hPEPT1, D2H and human  
 CC sucrose-isomaltase complex (hsi). (I) provide active transport of  
 CC therapeutic agents through human and animal GI tissue (into the blood)  
 CC for in vivo delivery, particularly for treatment or prevention  
 CC of hypertension, diabetes, osteoporosis, haemophilia, anaemia, cancer,  
 CC migraine, or angina pectoris. Specifically they are used to deliver  
 CC insulin or leuprolide, but many other suitable therapeutic agents are  
 CC disclosed, including genes or inhibitory nucleic acid, imaging agents and  
 CC antigens. (I) may also provide targeting to the GI tract. Other uses of  
 CC (I) are: (i) to determine the level of specified receptors in a sample  
 CC (in a binding assay); and (ii) to screen for molecules that bind (I); ...  
 CC immunogenic analogues or derivatives of (I) are used to raise antibodies  
 CC and in immunoassays. The antibodies are used to locate, detect and  
 CC measure (I), e.g. for imaging, monitoring treatment, tissue analysis  
 CC etc., also for peptide purification and immobilisation.  
 SQ Sequence 832 AA;

DT 17-APR-1995 (first entry)  
 DE Rat-224 cadherin partial sequence.  
 KW Cadherin; cell adhesion molecule.  
 OS Rattus rattus.  
 PN WO9414960-A.  
 PD 07-JUL-1994.  
 PR 23-DEC-1993; U12588.  
 PR 29-DEC-1992; US-998003.  
 PA (DOHE-) DOHENT EYE INST.  
 PI Suzuki S;  
 DR WPI: 94-293849/36.  
 DR N-PSDB: Q68957.  
 PT Polynucleotide sequences encoding new proto:cadherins - useful  
 PT for modulating natural binding and regulating activities.  
 PS Example, page 38, 114pp; English.  
 CC Two regions of conserved AA sequence, one from the middle of the  
 CC third cadherin extracellular subdomain (EC-3) and the other from the  
 CC C-terminus of the fourth extracellular subdomain (EC-4) were  
 CC identified. The corresp. degenerate oligos (Q68949, Q68950) were  
 CC designed for use as PCR primers. PCR was carried out on a rat brain  
 CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.  
 CC The 450 bp band corresponded to the expected length between the two  
 CC primer sites, but the 130 bp band could not be predicted from any  
 CC of the previously identified cadherin sequences. The 450 bp and 130  
 CC bp bands were extracted and sequenced. Nineteen novel partial cDNA  
 CC clones were isolated. The DNA and deduced AA sequences of the  
 CC clones (including sequences corresp. to the PCR primers) are given  
 CC in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the  
 CC cDNA clones are homologous to, but distinct from the known  
 CC cadherins. The cadherins described, thus far have highly conserved  
 CC short AA sequences in the EC-3 including the consensus sequence  
 CC D-Y-E or D-E-F located at the middle region of the subdomain and  
 CC the consensus sequence in R58879 or R58880 at its end, while the  
 CC corresp. sequences of other subdomains, except for the 5th extra-  
 CC cellular subdomain (EC-5), are D-R-E and the sequence in R58881  
 CC respectively. In contrast the deduced AA sequences of the new  
 CC clones that corresp. to cadherin extracellular subdomains include  
 CC the sequence D-Y-E or D-F-E at one end, but have the other end  
 CC D-X-N-D-N-X-P-X-F instead of R58879 or R58880 at the other end.  
 CC The polypeptides encoded by the partial clones are homologous to  
 CC previously identified cadherins but did not show significant  
 CC homology to any other sequences in Genbank. Therefore, the partial  
 CC cDNAs appear to comprise a new subclass of cadherin-related  
 CC molecules.  
 SQ Sequence 43 AA;  
 Query Match 56.8%; Score 50; DB 1; Length 43;  
 Best Local Similarity 66.7%; Pred. No. 0.085;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 CECNIRKVDVNDNFP 15  
 DB 29 CIIIRKVDVNDNFP 43  
 RESULT 8  
 R87107  
 ID R87107 standard; Peptide: 43 AA.  
 AC R87107;  
 DT 28-AUG-1996 (first entry)  
 DE Protocadherin clone RAT-224.  
 KW Protocadherin; pc3; pc4; pc5; human; rat; cadherin; cell adhesion;  
 KW catenin; therapy; clone.  
 OS Rattus rattus.  
 PN WO9600289-A1.  
 PD 04-JAN-1996.  
 PR 26-JUN-1995; U08071.  
 PR 27-JUN-1994; US-268161.  
 PA (DOHE-) DOHENT EYE INST.  
 PI Suzuki S;  
 DR WPI: 96-068873/07.  
 DR N-PSDB: T03582.

PT Polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat  
 PT pc5 - involved in cell-cell adhesion and regulation activities  
 PS Example 1; Page 40; 146pp; English.  
 CC R87102-R87120 represent partial fragments of the rat protocadherin  
 CC sequence. The cDNAs encoding these sequences were isolated after  
 CC screening a rat brain cDNA preparation with the primers shown in T03575  
 CC and T03576. The primers were constructed from portions of the amino acid  
 CC sequences of the third and fourth extracellular domains of published  
 CC cadherin sequences. The full length cDNA sequence encoding rat  
 CC protocadherin pc5 is represented in T03574. The cytoplasmic domain of  
 CC cadherin interacts with the cytoskeleton through catenins and other  
 CC cytoskeleton associated proteins. The cytoplasmic domain is not present  
 CC in all cadherins, but in those which possess it, it is essential for the  
 CC cadherin adhesive function. The cadherins which do not possess a  
 CC cytoplasmic domain appear to function via a different method from those  
 CC with a cytoplasmic domain. These protein sequences are involved in  
 CC cell-cell adhesion. These sequences may have regulatory functions in the  
 CC cell, as well as the cell-cell adhesive properties. Antipodies produced  
 CC against these sequences are useful for modulating the binding activity of  
 CC these protocadherins, and can be used therapeutically.  
 SO Sequence 43 AA;

Query Match 56.8%; Score 50; DB 1; Length 43;  
 Best Local Similarity 66.7%; Pred. No. 0.085;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 CECNIKVKVDNDFP 15  
 | : | | | | | | | | | |  
 Db 29 CKRIIKVVDNDNP 43

RESULT 9  
 W61273  
 ID W61273 standard; Protein: 1180 AA.  
 AC W61273;  
 DE Homo sapiens protocadherin-4.  
 KW pcch-4 gene; VE-cadherin; protocadherin-4; VE-cadherin-2;  
 KW vascular endothelial; angiogenesis; modulation; glioma; psoriasis;  
 KW inflammatory disease; organ transplantation; treatment; inhibition;  
 KW tumour; metastasis; rheumatoid arthritis; diagnosis; detection;  
 KW cell adhesion; atherosclerosis; myocardial ischemia.  
 OS Homo sapiens.  
 PN WO9825946-A1.  
 PD 18-JUN-1998.  
 PF 05-NOV-1997; U20006.  
 PR 12-DEC-1996; SE-004731.  
 PA (IMCL-) IMCLONE SYSTEMS INC.  
 PI Dejana E; Telo P;  
 DR WPI: 98-348441/30.  
 DR N-PSDB: V27593.  
 PT New isolated vascular endothelial cadherin-2 - used to develop  
 PT products for modulating angiogenesis, e.g. for treating tumours,  
 PT glioma, psoriasis, inflammatory diseases or organ transplantation  
 PS Claim 1: Page 31-37; 54pp; English.  
 CC The sequence is that of protocadherin-4 (pcch-4), or as it may  
 CC alternatively be known, VE-cadherin-2. As a vascular endothelial  
 CC cadherin it promotes cell-to-cell homotypic adhesion and its  
 CC expression is upregulated in proliferating endothelial cells in  
 CC comparison to resting cells. It can be used for inhibiting  
 CC angiogenesis and inhibiting pathological conditions such as tumours,  
 CC neovascular glioma, proliferative retinopathy including proliferative  
 CC diabetic retinopathy, macular degeneration, hemangiomas, angiodiomas,  
 CC and psoriasis. It may also be used for the prevention or inhibition of  
 CC leukocyte infiltration, tumour cell metastasis, or endothelial  
 CC permeability, as a vaccine and for making endothelial junctions more  
 CC permeable to antigens, thus indicating use of the modifiers for  
 CC treatment or prevention of acute and chronic inflammatory diseases,  
 CC organ transplantation, myocardial ischemia, atherosclerosis, rheumatoid  
 CC arthritis and intestinal infection. It may also be used for detection,  
 CC diagnosis and drug screening.  
 SO Sequence 1180 AA;

Query Match 56.8%; Score 50; DB 1; Length 1180;  
 Best Local Similarity 66.7%; Pred. No. 3.3;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 CECNIKVKVDNDFP 15  
 | : | | | | | | | | | |  
 Db 336 CKVLAQVLDYNDNP 350

RESULT 10  
 R58875  
 ID R58875 standard; Protein: 41 AA.  
 AC R58875;  
 DE Rat-352 cadherin partial sequence.  
 DE Rat-352 cadherin partial sequence.  
 KW Cadherin; cell adhesion molecule.  
 OS Rattus rattus.  
 PN WO9414867-A1.  
 PD 07-JUL-1994.  
 PF 23-DEC-1993; U12588.  
 PR 29-DEC-1992; US-998003.  
 PA (DOHE-) DOHERN EYE INST.  
 PI Suzuki S;  
 DR WPI: 94-293849/36.  
 DR N-PSDB: Q68967.

PT Polynucleotide sequences encoding new proto:cadherins - useful  
 PT for modulating natural binding and regulating activities.  
 PS Example; page 44; 114pp; English.  
 CC Two regions of conserved AA sequence, one from the middle of the  
 CC third cadherin extracellular subdomain (EC-3) and the other from the  
 CC C-terminus of the fourth extracellular subdomain (EC-4) were  
 CC identified. The corresp. degenerate oligos (Q68949, Q68950) were  
 CC designed for use as PCR primers. PCR was carried out on a rat brain  
 CC cDNA prep. Two major bands of about 450 bps and 130 bps were found.  
 CC The 450 bp band corresponded to the expected length between the two  
 CC primer sites, but the 130 bp band could not be predicted from any  
 CC of the previously identified cadherin sequences. The 450 bp and 130  
 CC bp bands were extracted and sequenced. Nineteen novel partial cDNA  
 CC clones were isolated. The DNA and deduced AA sequences of the  
 CC clones (including sequences corresp. to the PCR primers) are given  
 CC in Q68951-Q68969 and R58860-R58878. The deduced AA sequences of the  
 CC cDNA clones are homologous to, but distinct from the known  
 CC cadherins. The cadherins described thus far have highly conserved  
 CC short AA sequences in the EC-3 including the consensus sequence  
 CC D-Y-E or D-E-F located at the middle region of the subdomain and  
 CC the consensus sequence in R58879 or R58880 at its end, while the  
 CC corresp. sequences of other subdomains, except for the 5th extra-  
 CC cellular subdomain (EC-5), are D-R-E and the sequence in R58881  
 CC respectively. In contrast the deduced AA sequences of the new  
 CC clones that corresp. to cadherin extracellular subdomains include  
 CC the sequence D-Y-E or D-E-F at one end, but have the sequence  
 CC D-X-N-N-X-P-X-F instead of R58879 or R58880 at the other end.  
 CC The polypeptides encoded by the partial clones are homologous to  
 CC previously identified cadherins but did not show significant  
 CC homology to any other sequences in Genbank. Therefore, the partial  
 CC cDNAs appear to comprise a new subclass of cadherin-related  
 CC molecules.  
 SO Sequence 41 AA;

Query Match 54.5%; Score 48; DB 1; Length 41;  
 Best Local Similarity 60.0%; Pred. No. 0.17;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 CECNIKVKVDNDFP 15  
 | : | | | | | | | | | |  
 Db 27 CTVSIQVVDNDNP 41

RESULT 11  
 R87117





DT 13-MAY-1997 (first entry)  
 DE Full length rat cadherin-8.  
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
 KW brain; human; antibody; purification; determination;  
 KW tissue expression; binding antagonist; calcium ion.  
 OS Rattus rattus.  
 PN US5597725-A.  
 PD 28-JAN-1997.  
 PF 17-APR-1992; 872643.  
 PR 17-APR-1992; US-872643.  
 PR 19-APR-1993; US-049460.  
 PR 26-JAN-1994; US-188228.  
 PA (DOHE-) DOHENY EYE INST.  
 PI SUZUKI S;  
 DR WPI: 97-108328/10.  
 DR N-PSDB: T61917.  
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
 PT etc.  
 PS Example 1; Columns 49-52; 59pp; English.  
 CC The present sequence is full length rat cadherin-8, which is a  
 CC Ca2+ dependent cell adhesion protein. The rat cadherin cDNA was  
 CC isolated from a brain cDNA library using a labelled rat cadherin-8  
 CC extracellular domain PCR fragment as a probe. The rat cDNA was then  
 CC used to isolate the cDNA encoding its full length human homologue,  
 CC from a human foetal brain cDNA library.  
 CC Antibodies or fragments that specifically bind the human cadherin  
 CC can be used to purify the cadherin, determine its tissue expression  
 CC and antagonise its ligand/antiligand binding activities.  
 SQ Sequence 799 AA;

Query Match 52.3%; Score 46; DB 1; Length 799;  
 Best Local Similarity 72.7%; Pred. No. 9.7;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 IKVKDVNDNFP 15  
 |||:|||||  
 DB 155 IKVODINDNAP 165

## RESULT 15

W13132  
 ID W13132 standard; Protein; 793 AA.  
 AC W13132;  
 DT 14-MAY-1997 (first entry)  
 DE Full length human cadherin-8.  
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
 KW brain; human; antibody; purification; determination;  
 KW tissue expression; binding antagonist; calcium ion.  
 OS Homo sapiens.  
 PN US5597725-A.  
 PD 28-JAN-1997.  
 PF 17-APR-1992; 872643.  
 PR 17-APR-1992; US-872643.  
 PR 19-APR-1993; US-049460.  
 PR 26-JAN-1994; US-188228.  
 PA (DOHE-) DOHENY EYE INST.  
 PI SUZUKI S;  
 DR WPI: 97-108328/10.  
 DR N-PSDB: T61923.  
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
 PT etc.  
 PS Example 2; Columns 85-90; 59pp; English.  
 CC The present sequence is full length human cadherin-8, which  
 CC is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA  
 CC was isolated from a foetal brain cDNA library, using probes based  
 CC on homologous rat cadherin cDNA.  
 CC Antibodies or fragments that specifically bind the human cadherin  
 CC can be used to purify the cadherin, determine its tissue expression  
 CC and antagonise its ligand/antiligand binding activities.  
 SQ Sequence 793 AA;

Query Match 52.3%; Score 46; DB 1; Length 793;  
 Best Local Similarity 72.7%; Pred. No. 9.6;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 OY 5 IKVKDVNDNFP 15  
 |||:|||||  
 DB 148 IKVODINDNAP 158

Search completed: January 6, 2000, 15:40:37  
 Job time: 1664 sec

Sat Jan 15 11:45:03 2000

us-08-991-628-5.rag

Page 7

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:37 : Search time 51.33 Seconds  
(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-6  
Perfect score: 82  
Sequence: 1 SARTLNRRYTPYTF 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A.Geneseq\_36.\*

Word size: 0

Number of hits that pass the threshold: 188963

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	100.0	999	1 R30742	Human pemphigus vu
2	82	100.0	614	1 W07908	Pemphigus vulgaris
3	82	100.0	15	1 W04846	Self epitope of de
4	38	46.3	431	1 R97293	Mouse CRF RBL rece
5	37	45.1	431	1 R90575	Rat CRF2-Beta rece
6	37	45.1	454	1 R93616	Kaposi's sarcoma a
7	37	45.1	454	1 R97842	N-acetylcholinesterase
8	36.5	44.5	820	1 R71802	Haem 84-1 portion
9	36	43.9	516	1 R12098	Haem 84-2 portion
10	36	43.9	480	1 R12099	Hog cholesta virus
11	36	43.9	3898	1 R10473	Bloom's syndrome act
12	36	43.9	1418	1 W15264	Bloom's syndrome B
13	36	43.9	185	1 W31545	Bloom's syndrome B
14	36	43.9	271	1 W31546	Bloom's syndrome B
15	36	43.9	515	1 W31547	Bloom's syndrome B
16	36	43.9	1417	1 W31548	Bloom's syndrome B
17	36	43.9	739	1 W31549	Bloom's syndrome B
18	36	43.9	1417	1 W31550	Bloom's syndrome B
19	36	43.9	1417	1 W31551	Bloom's syndrome B
20	36	43.9	244	1 W77619	Mercuric reductase
21	36	43.9	344	1 W98545	H. pylori GHP0.156
22	36	43.9	1261	1 W93601	Tomato Ka21 clone
23	35	42.7	325	1 R35522	Protein kinase (CK
24	35	42.7	364	1 R56523	Protein kinase (CK
25	35	42.7	337	1 R56524	Protein kinase (CK
26	35	42.7	369	1 R62443	R. clyti encoded by
27	35	42.7	364	1 R76619	Human HRR25-like C
28	35	42.7	337	1 R76620	Human HRR25-like C
29	35	42.7	325	1 R76618	Human HRR25-like C
30	35	42.7	1196	1 R75189	Osteolinductin ret
31	35	42.7	693	1 R88646	Neisseria meningit
32	35	42.7	610	1 R88654	N.meningitidis 694
33	35	42.7	752	1 R97199	Chimeric thymost
34	35	42.7	429	1 W08606	Chimeric thymost
35	35	42.7	449	1 W08608	Chimeric thymost
36	35	42.7	428	1 W08609	Chimeric thymost
37	35	42.7	105	1 W52243	Antibody LDI/2-6-3
38	35	42.7	105	1 W52241	Antibody LDI/2-6-3
39	35	42.7	933	1 W98599	H. pylori GHP0.127

## ALIGNMENTS

40	35	42.7	623	1 W30620	Arabidopsis La-er
41	35	42.7	623	1 W30622	Arabidopsis enhanc
42	35	42.7	623	1 W30623	Arabidopsis enhanc
43	35	42.7	623	1 W30624	Arabidopsis mutant
44	35	42.7	1079	1 R03163	MUV reverse trans
45	34	41.5	354	1 R42424	Rat gustducin alph

  

RESULT 1	1	100.0%	Score 82; DB 1; Length 999;
R30742	1	100.0%	Best Local Similarity 100.0%; Pred. No. 2.1e-06;
R30742	1	100.0%	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB	512 SARTLNRRYTPYTF 526		

  

Query Match	100.0%	Score 82; DB 1; Length 999;
Best Local Similarity	100.0%; Pred. No. 2.1e-06;	
Matches 15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 SARTLNRRYTPYTF 15	
DB	512 SARTLNRRYTPYTF 526	

  

RESULT 2	2	100.0%	Score 82; DB 1; Length 999;
W07908	2	100.0%	Best Local Similarity 100.0%; Pred. No. 2.1e-06;
W07908	2	100.0%	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
ID	W07908 standard; protein; 614 AA.		
AC	W07908: 29-JAN-1997 (first entry)		
DE	Pemphigus vulgaris antigen protein extracellular region.		
KW	Anticardiolipin; immunoglobulin G; IgG1; fusion protein; diagnosis; treatment; pemphigus vulgaris; PV; bulla; blister; skin disease; dermatology.		
KW	Homo sapiens.		
OS	Homo sapiens.		
PN	J08188540-A.		
PD	23-JUL-1996.		
PE	30-JUN-1995; 165632.		
PR	30-JUN-1994; JP-173291.		
PA	(NISH/) NISHIKAWA T.		
DR	WPI; 96-388562/38.		
PT	Fused protein recognised by pemphigus vulgaris autoantibody -		
PS	useful to treat and diagnose pemphigus vulgaris		
CC	Claim 1; Page 7-9; 9PP; Japanese.		
CC	W07908 represents the human pemphigus vulgaris (PV) antigen		
CC	extracellular region. The PV antigen is produced in patients with		
CC	pemphigus vulgaris resulting in autoimmune disease. PV is a rare		
CC	relapsing disease-causing suprabasal, intra-epidermal bullae		
CC	(vesicles) of the skin and mucous membranes, which is fatal if		
CC	untreated. The PV antigen was fused to a human IgG1 hinge region		
CC	and the resulting fusion protein is useful to treat or diagnose		

CC pemphigus vulgaris.  
SQ Sequence 614 AA;

Query Match 100.0%; Score 82; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNRYTGPYTF 15  
Db 511 SARTLNRYTGPYTF 525

RESULT 3  
W04846  
ID W04846 standard; peptide; 15 AA.  
AC W04846;  
DT 18-FEB-1997 (first entry)  
DE Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerisation; self-epitope; antigen; autoimmune disease;  
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphomannomutase;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW influenza; haemagglutinin; reovirus; sigma protein.  
KM Homo sapiens.  
PN W09627387-A1.  
PD 12-SEP-1996.  
PE 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PA (HARD) HARVARD COLLEGE.  
PI Strominger JL, Mocherperennig KW;  
DR WPI; 96-425218/42.  
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for  
PT identification of other self and non-self antigens implicated in  
PT autoimmune disease  
PS Claim 1; Page 41; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise  
CC either an isolated human non-collagen or non-mysin basic protein  
CC (MBP) polypeptide which is capable of tolerising an individual to an  
CC autoantigen; or an isolated human pathogen polypeptide capable of  
CC tolerising an individual to that polypeptide. In both cases, the  
CC polypeptide (whether self or non-self) includes an amino acid  
CC sequence corresponding to a sequence motif for a MHC class II  
CC protein, such as HLA-DR, which is associated with a human autoimmune  
CC disease and which binds to the polypeptide to activate autoreactive  
CC T-cells in individuals with the autoimmune disease. This peptide is  
CC derived from the human desmoglein 3 protein (amino acids 512-526)  
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
CC derived from the human desmoglein protein are described in W04841-47.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 82; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.9e-08;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNRYTGPYTF 15  
Db 1 SARTLNRYTGPYTF 15

RESULT 4  
R97293  
ID R97293 standard; Protein; 431 AA.  
AC R97293;  
DT 21-AUG-1996 (first entry)  
DE Mouse CRF Rb1 receptor.  
KW Corticotropin releasing factor receptor; CRF-R; corticoliberin;  
KW signal transduction.  
OS Mus sp.  
FH Key

Location/Qualifiers

FT modified\_site 52  
FT /label- N-glycosylation\_site  
FT modified\_site 61  
FT /label- N-glycosylation\_site  
FT modified\_site 94  
FT /label- N-glycosylation\_site  
FT modified\_site 105  
FT /label- N-glycosylation\_site  
FT modified\_site 113  
FT /label- N-glycosylation\_site  
FT domain 139.159  
FT /label- Transmembrane\_domain-1  
FT domain 169.188  
FT /label- Transmembrane\_domain-2  
FT domain 206.229  
FT /label- Transmembrane\_domain-3  
FT domain 245.265  
FT /label- Transmembrane\_domain-4  
FT domain 285.307  
FT /label- Transmembrane\_domain-5  
FT domain 331.351  
FT /label- Transmembrane\_domain-6  
FT domain 366.385  
FT /label- Transmembrane\_domain-7  
PN W09617934-A2.  
PD 13-JUN-1996.  
PE 06-DEC-1995; U15909.  
PR 09-DEC-1994; US-353537.  
PR 17-JAN-1995; US-374009.  
PA (SALK) SALK INST BIOLOGICAL STUDIES.  
PI Chen R, Donaldson CJ, Lewis KA, Perrin MH, Sawchenko P;  
PI Vale WW; 96-287179/29.  
DR N-PSDB; T28972.  
DT WPI; 96-287179/29.  
PT Isolated corticotropin-releasing factor receptor (CRF-R) - used to  
PT develop prods. for modulating signal transduction activity mediated  
PT by CRF-R  
PS Claim 3; Page 83-85; 102pp; English.  
CC Mouse corticotropin releasing factor receptor mCRF-Rb1 was  
CC identified as the product of a cDNA clone (728972) isolated from a  
CC mouse heart library. Recombinant mCRF-Rb1 can be expressed in  
CC host cells transformed by the cDNA clone. The receptor can be used  
CC to identify agonists and antagonists that modulate the signal  
CC transduction activity mediated by CRF receptors. It may be  
CC administered therapeutically to reduce high ACTH levels caused by  
CC excess CRF.  
SQ Sequence 431 AA;

Query Match 46.3%; Score 36; DB 1; Length 431;  
Best Local Similarity 46.2%; Pred. No. 49;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 RTLNRYTGPYTF 15  
Db 47 RTLNRYTGPYTF 59

RESULT 5  
R90575  
ID R90575 standard; Protein; 431 AA.  
AC R90575;  
DT 08-APR-1996 (first entry)  
DE Rat CRF2-beta receptor.  
KW CRF2-beta receptor; corticotropin-releasing factor-2 receptor;  
KW cerebrovascular disorder; memory disorder; Alzheimer disease.  
OS Rattus sp.  
FH Key  
FH Location/Qualifiers  
FT domain 1.117  
FT /label- Extracellular\_N-terminal\_domain  
FT domain 118.138  
FT /label- Transmembrane\_domain  
FT domain 139.147

FT domain /label= Intracellular\_domain  
 FT domain 148..167 /label= Transmembrane\_domain  
 FT domain 168..184 /label= Transmembrane\_domain  
 FT domain 185..208 /label= Extracellular\_domain  
 FT domain 209..223 /label= Transmembrane\_domain  
 FT domain 224..244 /label= Intracellular\_domain  
 FT domain 245..261 /label= Transmembrane\_domain  
 FT domain 262..286 /label= Extracellular\_domain  
 FT domain 287..309 /label= Transmembrane\_domain  
 FT domain 310..329 /label= Intracellular\_domain  
 FT domain 330..342 /label= Transmembrane\_domain  
 FT domain 343..363 /label= Extracellular\_domain  
 FT domain 364..431 /label= Transmembrane\_domain  
 FT domain /label= C-terminal\_intracellular\_domain  
 FT W09534651-A2.  
 PN 21-DEC-1995.  
 PD 14-JUN-1995: U07757.  
 PR 14-JUN-1994: US-259959.  
 PR 31-JAN-1995: US-381433.  
 PR 07-JUN-1995: US-485984.  
 PA (NEUR-) NEUROCRINE BIOSCIENCES INC.  
 PI Chambers D, De Souza EB, Grigoriadis DE, Llaw CW;  
 PI Lovenberg TW, Oltersdorf T;  
 DR WPI: 96-049680/05.  
 DR N-PSDB: T12244.  
 PT Corticotropin-releasing factor-2 receptor, and DNA encoding it  
 PT used to isolate CRF-2 receptor antagonists for the treatment of  
 PT cerebrovascular disorders, memory disorders and Alzheimer's disease  
 PT Disclosure: Page 63-66; 109pp; English.  
 CC Rat corticotropin-releasing factor-2-beta (CRF2-beta) receptor  
 CC (R90575) is a membrane-bound G-coupled protein involved  
 CC in signal transduction. It can be produced by expression of  
 CC encoding cDNA (T12244) in prokaryotic or eucaryotic host cells.  
 CC Recombinant CRF2 receptor is used to screen CRF2 receptor agonists  
 CC and antagonists of therapeutic appln., and to prepare antibodies  
 CC which specifically bind to CRF2 receptors.  
 SQ Sequence 431 AA;

Query Match 45.1%; Score 37; DB 1; Length 431;  
 Best Local Similarity 38.5%; Pred. No. 73;  
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 RTLNRYTGP 15  
 II : : : : :  
 Db 47 RTTRNFSGPYSY 59

RESULT 6  
 R93616  
 ID R93616 standard; Protein; 454 AA.  
 AC R93616;  
 DT 13-AUG-1996 (first entry)  
 DE Kaposi's sarcoma associated herpesvirus ORF32 product.  
 KW Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis;  
 KW therapy.  
 OS Kaposi's sarcoma associated herpesvirus.  
 PN W09606159-A1.  
 PD 29-FEB-1996.  
 PE 11-AUG-1995: U10194.  
 PR 18-AUG-1994: US-292365.  
 PR 21-NOV-1994: US-343100.

PR 11-APR-1995: US-420235.  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 PA (GRAN/) GRANT D E.  
 PA (VIEL/) VIELE L.  
 PI Chang Y, Moore PS, Grant DE, Vile L;  
 DR WPI: 96-151362/15.  
 DR N-PSDB: T16822.  
 PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated  
 PT vectors and proteins, used in detection and vaccination.  
 PS Claim 17; Page 235-237; 305pp; English.  
 CC Kaposi's sarcoma associated herpes virus (KSHV) clone KS5 (T16806),  
 CC obd. from a KS lesion genomic library, includes 15 complete ORFs and  
 CC 2 incomplete ORFs (T16807-23) named according to their herpesvirus  
 CC salmt positional homologues. The ORF32 (T16822) product is  
 CC given in R93616. KSHV proteins and peptides may be obd. by  
 CC incorporating encoding sequences into a vector and expression in  
 CC host cells. They are useful in vaccines or for raising antibodies of  
 CC diagnostic or therapeutic value.  
 SQ Sequence 454 AA;

Query Match 45.1%; Score 37; DB 1; Length 454;  
 Best Local Similarity 54.5%; Pred. No. 77;  
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 ARTLNRYTGP 12  
 I : : : : :  
 Db 3 AHAINERYVGP 13

RESULT 7  
 R97842  
 ID R97842 standard; Protein; 454 AA.  
 AC R97842;  
 DT 11-SEP-1996 (first entry)  
 DE Kaposi's sarcoma associated herpesvirus ORF32 product.  
 KW Kaposi's sarcoma; gamma-2 herpesvirus; KSHV; therapy; diagnosis;  
 KW vaccine; diagnosis; AIDS.  
 OS Kaposi's sarcoma associated herpesvirus.  
 PN W09615779-A1.  
 PD 30-MAY-1996.  
 PE 21-NOV-1995: U15138.  
 PR 21-NOV-1994: US-343101.  
 PR 11-APR-1995: US-420235.  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 PI Chang Y, Moore PS;  
 PI WPI: 96-268320/27.  
 DR N-PSDB: T30694.  
 PT Herpes virus associated with Kaposi's sarcoma - also definitive DNA  
 PT sequences, useful for diagnosis of and to develop prods. for  
 PT treatment of Kaposi's sarcoma  
 PS Claim 17; Page 188-189; 277pp; English.  
 CC Lambda clone KS5 (T30681) is a fragment of a newly identified human  
 CC gamma-2 herpesvirus associated with Kaposi's sarcoma (KS). KS5 has  
 CC 17 open reading frames (T30682-98), 15 of which are complete,  
 CC including ORF32 (T30694). The protein products (R97830-46,  
 CC respectively) of the 17 ORFs can be expressed in eukaryotic or  
 CC bacterial host cells for use as vaccines, for KS diagnosis, or for  
 CC raising antibodies.  
 SQ Sequence 454 AA;

Query Match 45.1%; Score 37; DB 1; Length 454;  
 Best Local Similarity 54.5%; Pred. No. 77;  
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 ARTLNRYTGP 12  
 I : : : : :  
 Db 3 AHAINERYVGP 13

RESULT 8  
 R71802

ID R71802 standard; Protein: 820 AA.  
 AC R71802; 1995 (first entry)  
 DT 19-OCT-1995  
 DE N-acetylheparosan fragmentation enzyme.  
 DE N-acetylheparosan; polysaccharide K5; fragmentation; degradation;  
 KW Pharmaceutical starting material.  
 OS Escherichia coli (K5) SEBR 3282 (CNCM I-1013).  
 FH Key location/Qualifiers  
 FT binding\_site 384..387  
 FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr  
 FT and X2-Val, Ile or Ala)"  
 FT binding\_site 411..414  
 FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr  
 FT and X2-Val, Ile or Ala)"  
 FT binding\_site 433..436  
 FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr  
 FT and X2-Val, Ile or Ala)"  
 FT binding\_site 461..464  
 FT /label= potential lipoprotein attachment site  
 FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr  
 FT and X2-Val, Ile or Ala)"  
 FT binding\_site 495..498  
 FT /note= "Asn-X1-X2-Ser motif (X1-Ile, Leu or Tyr  
 FT and X2-Val, Ile or Ala)"  
 FT region 804..809  
 FT /note= "strongly charged C-terminal region"  
 FT  
 PN FR2709132-A.  
 PD 24-FEB-1995.  
 PF 17-AUG-1993; 010050.  
 PR 17-AUG-1993; FR-010050.  
 PA (SNFI) ELF SANOFI.  
 PI Legoux R, Lelong P, Salome MLV;  
 DR N-PSDB: 086268.  
 DT New DNA encoding enzyme for fragmenting N-acetylheparosan -  
 PT giving products useful as pharmaceutical starting material, also  
 PT related protein, vector transformed cells, etc.  
 PS Claim 2: Page 38-40; 54pp; French.  
 CC An enzyme which degrades high mol. wt. N-acetylheparosan into  
 CC fragments of lower mol. wt. (more suitable for use as pharmaceutical  
 CC starting materials) can be obtained from E. coli (K5) SEBR 3282. The  
 CC DNA sequence (086268) coding for the enzyme was isolated in a  
 CC plasmid (p838.7) which has been deposited in E. coli RFL strain as  
 CC CNCM I-1352. The deduced amino acid sequence (R71802) has homology  
 CC with an exo-poly-alpha-D-galacturonidase from Erythrina chrysanthem.  
 SQ Sequence 820 AA;

Query Match 44.5%; Score 36.5; DB 1; Length 820;  
 Best Local Similarity 47.1%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 1; Mismatches 1; Indels 7; Gaps 1;

OY 6 LNNRYTG-----PYTF 15  
 : ||| :  
 Db 9 HNEITGNGVTISPYTF 25

RESULT 9  
 ID R12098  
 ID R12098 standard; Protein: 516 AA.  
 AC R12098;  
 DT 22-JUL-1991 (first entry)  
 DE Haem 84-1 portion of H. contortus 55A surface protein.  
 KW Parasitic nematode; vaccine; Hc.  
 OS Haemonchus contortus.  
 PN AU9062569-A.  
 PD 21-MAR-1991.  
 PF 17-SEP-1990; 062569.  
 PR 18-SEP-1989; US-408339.  
 PR 01-MAR-1990; US-487181.  
 PA (SYNE-) SYNERGEN INC.  
 DR WPI: 91-133285/19.  
 DR N-PSDB: Q11798.

PT Protein from Haemonchus contortus and other nematodes - used as  
 PT therapeutic and prophylactic agent to protect plants, animals or  
 PT humans from parasitic nematode infection.  
 PS Disclosure: Fig 49; 209pp; English.  
 CC The proteins derived from the nematode DNA may be used to derive  
 CC vaccines against parasitic infection of plants, humans and animals  
 CC esp. sheep. MAb may also be raised to provide passive therapy and  
 CC prophylaxis against infection.  
 SQ Sequence 516 AA;

Query Match 43.9%; Score 36; DB 1; Length 516;  
 Best Local Similarity 54.5%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 LNNRYTGPTTF 15  
 : ||| :  
 Db 484 VNNRYGGRXY 494

RESULT 10  
 ID R12099  
 ID R12099 standard; Protein: 480 AA.  
 AC R12099;  
 DT 22-JUL-1991 (first entry)  
 DE Haem 84-2 portion of H. contortus 55A surface protein.  
 KW Parasitic nematode; vaccine; Hc.  
 OS Haemonchus contortus.  
 PN AU9062569-A.  
 PD 21-MAR-1991.  
 PF 17-SEP-1990; 062569.  
 PR 18-SEP-1989; US-408339.  
 PR 01-MAR-1990; US-487181.  
 PA (SYNE-) SYNERGEN INC.  
 DR WPI: 91-133285/19.  
 DT New DNA encoding enzyme for fragmenting N-acetylheparosan -  
 PT giving products useful as pharmaceutical starting material, also  
 PT related protein, vector transformed cells, etc.  
 PS Claim 2: Page 38-40; 54pp; French.  
 CC An enzyme which degrades high mol. wt. N-acetylheparosan into  
 CC fragments of lower mol. wt. (more suitable for use as pharmaceutical  
 CC starting materials) can be obtained from E. coli (K5) SEBR 3282. The  
 CC DNA sequence (086268) coding for the enzyme was isolated in a  
 CC plasmid (p838.7) which has been deposited in E. coli RFL strain as  
 CC CNCM I-1352. The deduced amino acid sequence (R71802) has homology  
 CC with an exo-poly-alpha-D-galacturonidase from Erythrina chrysanthem.  
 SQ Sequence 480 AA;

Query Match 43.9%; Score 36; DB 1; Length 480;  
 Best Local Similarity 54.5%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 LNNRYTGPTTF 15  
 : ||| :  
 Db 376 VNNRYGGRXY 386

RESULT 11  
 ID R10473  
 ID R10473 standard; Protein: 3898 AA.  
 AC R10473;  
 DT 09-APR-1991 (first entry)  
 DE Hog cholera virus genome product.  
 KW Swine fever virus; HCV; pestivirus; border disease virus; BDV;  
 KW Bovine viral diarrhoea virus; BVDV; Togaviridae; ss.  
 OS Hog cholera virus.  
 PN WO9100352-A.  
 PD 10-JAN-1991.  
 PF 29-JUN-1989; NL-001651.  
 PR 29-JUN-1989; NL-001651.  
 PA (DIER-) CENT DIERGENESKUND.  
 PI Moorman R, Mensvoort G;  
 DR WPI: 91-036746/05.  
 DR P-PSDB: Q10353.

PT Protection of animals against microbial infection - using  
 PT nucleotide sequences related to the microbe and a stop transfer  
 PT region  
 PS Clam 15; Fig 1; 33pp; English.  
 CC The genome sequence and its encoded product are sufficiently similar  
 CC to other pestiviruses of the family Flaviviridae, that they may be  
 CC used in vaccination and diagnosis of pestivirus infection.  
 SQ Sequence 3898 AA;

Query Match 43.9%; Score 36; DB 1; Length 3898;  
 Best Local Similarity 63.6%; Pred. No. 1.3e+03;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 2 ARTLNRYTGP 12  
 DB 988 AKTLRNRYPPE 998

## RESULT 12

W15264  
 ID W15264 standard; Protein; 1418 AA.  
 AC W15264;  
 DT 26-JAN-1998 (first entry)  
 DE Bloom syndrome active BLM protein.  
 KW BLM: Bloom syndrome; BS; mutant; probe: PCR primer; cancer;  
 KW therapy; diagnosis; vector.  
 OS Homo sapiens.  
 PN MO9717979-A1.  
 PD 22-MAY-1997.  
 PE 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR N-PSDB: T67013.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM  
 PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Clam 58; Fig 2; 51pp; English.  
 CC This active BLM protein is encoded by a 4437 bp BLM gene sequence.  
 CC This is used in diagnosing Bloom's syndrome where 2 mutated BLM genes  
 CC or the absence of a wild-type BLM gene in the nucleic acid of a subject  
 CC is detected. The BLM gene is in the 250 kb region between two markers  
 CC binding a 1.3 cm region on chromosome 15. cDNA from this region was  
 CC isolated, a 847 bp fragment selected and used to screen a HeLa cDNA  
 CC library. The longest clone H1 was isolated and extended by PCR to isolate  
 CC this BLM gene sequence. Delivery of a functional BLM gene to bone marrow  
 CC cells, e.g. via the viral vectors, is used to treat or prevent the onset  
 CC of Bloom's syndrome. Identification of the BLM gene and its products  
 CC should assist in the development of therapeutic and diagnostic agents for  
 CC cancer.  
 SQ Sequence 1418 AA;

Query Match 43.9%; Score 36; DB 1; Length 1418;  
 Best Local Similarity 87.5%; Pred. No. 4.2e+02;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNNR 8  
 DB 17 SARTLNKK 24

## RESULT 13

W15145  
 ID W15145 standard; Protein; 165 AA.  
 AC W15145;  
 DT 27-JAN-1998 (first entry)  
 DE Bloom's syndrome BLM mutated protein.  
 KW BLM: Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers

FT Protein 1.185  
 FT /note="Truncated protein. The deletion of CAA at  
 FT nucleotide positions 631-633 of the wild  
 FT type BLM gene results in a stop codon at  
 FT amino acid position 186"

MO9717979-A1.  
 PD 22-MAY-1997.  
 PE 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR N-PSDB: T93389.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM  
 PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Disclosure; Page -; 51pp; English.  
 CC This is a truncated BLM protein encoded by a mutant gene sequence  
 CC isolated from a Japanese Bloom's Syndrome sufferer designated  
 CC "97(ASOK)". The deletion of CAA from positions 631-633 of the  
 CC wild-type gene results in a stop codon at amino acid position 186.  
 CC This was one of the seven unique mutations which were identified in  
 CC a study of 10 people with Bloom's Syndrome. Based on the various  
 CC mutations, diagnostic tests for Bloom's Syndrome have been developed  
 CC which use standard sequence analysis techniques to detect the presence  
 CC of 2 mutated BLM genes or the absence of a wild-type BLM gene.  
 CC Note: The present sequence does not appear in the specification. It  
 CC has been made by modifying the H1-5' wild type BLM sequence which is  
 CC provided in Figure 2 (W15264).  
 SQ Sequence 185 AA;

Query Match 43.9%; Score 36; DB 1; Length 185;  
 Best Local Similarity 87.5%; Pred. No. 42;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNNR 8  
 DB 17 SARTLNKK 24

## RESULT 14

W15146  
 ID W15146 standard; Protein; 271 AA.  
 AC W15146;  
 DT 27-JAN-1998 (first entry)  
 DE Bloom's syndrome BLM mutated protein.  
 KW BLM: Bloom's syndrome; BS; mutant; truncation; therapy; diagnosis.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Protein 1.271  
 FT /note="Truncated protein. The substitution of the base A  
 FT at the position 888 of the wild type H1-5' gene  
 FT to the base T results in a stop codon at amino  
 FT acid position 272"  
 MO9717979-A1.  
 PD 22-MAY-1997.  
 PE 15-NOV-1996; U19046.  
 PR 15-NOV-1995; US-559303.  
 PA (NYBL-) NEW YORK BLOOD CENT INC.  
 PI Ellis N, German J, Groden J;  
 DR WPI: 97-289051/26.  
 DR N-PSDB: T93390.  
 PT Diagnosing Bloom's syndrome, and carriers, by detecting mutant BLM  
 PT genes - for gene therapy with nucleic acid encoding active BLM  
 PT protein to treat Bloom's syndrome and cancer in general  
 PS Disclosure; Page -; 51pp; English.  
 CC This is a truncated BLM protein encoded by a mutant gene sequence  
 CC isolated from a German Bloom's Syndrome sufferer designated "112(Nasch)".  
 CC The substitution of the base A at the position 888 of the wild type H1-5'  
 CC gene to the base T results in a stop codon at amino acid position 272.  
 CC This was one of the seven unique mutations which were identified in  
 CC a study of 10 people with Bloom's Syndrome. Based on the various





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:44 ; Search time 65.16 Seconds  
(without alignments)  
10.858 Million cell updates/sec

Title: US-08-991-628-6  
Perfect score: 82  
Sequence: 1 SARTLNRRYGPYTF 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR\_62:\*

Word size: 0

Number of hits that pass the threshold: 142080

1: PIR1:\*\n2: PIR2:\*\n3: PIR3:\*\n4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	100.0	999	1 IJHUG3	desmoglein 3 precu
2	46	56.1	531	2 C70333	hypothetical prote
3	42	51.2	163	3 B70854	hypothetical prote
4	40	48.8	848	4 A44282	retrovirus-related
5	39	47.6	2261	1 A42548	genome polypotein
6	39	47.6	2261	1 A42548	large protein L -
7	39	47.6	282	2 S43577	C28A5.4 protein (c
8	39	47.6	973	2 T01862	hypothetical prote
9	39	47.6	431	2 S25196	mitochondrial impo
10	38	46.3	865	2 S69044	hypothetical prote
11	38	46.3	431	2 I49279	saavagine/corticot
12	38	46.3	336	2 F71151	probable glucose-1
13	38	46.3	244	2 C70912	hypothetical prote
14	38	46.3	423	2 B70931	probable PPE prote
15	38	46.3	433	2 E69312	nitrate reductase
16	38	46.3	430	2 S46726	corticoidliperin rec
17	38	46.3	431	2 I49149	CRF receptor - mou
18	38	46.3	332	2 A75098	glucose-1-phosphat
19	37.5	45.7	690	2 T01183	hypothetical prote
20	37	45.1	387	2 A45117	aspartic proteinas
21	37	45.1	1186	2 S70430	hypothetical prote
22	37	45.1	368	2 A69275	3-ketosecy-CoA thi
23	37	45.1	302	2 F64883	probable transcript
24	37	45.1	218	2 H68005	hypothetical prote
25	37	45.1	405	2 A69206	hypothetical prote
26	37	45.1	413	2 S28066	esci protein - fls
27	37	45.1	643	2 S46723	arginine--tRNA lig
28	37	45.1	492	2 S54536	probable membrane
29	37	45.1	331	2 T75067	aminopeptidase fro
30	36.5	44.5	980	2 T00045	cellohextrin phosp
31	36	43.9	2269	1 J01750	genome polypotein
32	36	43.9	362	1 VVVP15	coat protein VP1 -
33	36	43.9	372	1 VVVPK1	coat protein VP1 -
34	36	43.9	3898	1 GNMVHB	genome polypotein
35	36	43.9	1237	1 GNFPA2	retrovirus-related

basepiate protein  
probable gpa2 pro  
NAH dehydrogenase  
probable diaphenol  
Bloom's syndrome r  
UDP-glucose 4-epim  
udp-glucose 4-epim  
coat protein VP1 -  
extensin homolog F  
hypothetical prote

## ALIGNMENTS

RESULT 1  
IJHUG3  
desmoglein 3 precursor - human  
N:Alternate names: pemphigus vulgaris antigen  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: A41088  
R:Amagali, M.; Klaus-Kovtun, V.; Stanley, J.R.  
Cell 67, 869-877, 1991  
A:Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a  
A:Reference number: A41088; MUID:92069753  
A:Accession: A41088  
A:Molecule type: mRNA  
A:Residues: 1-999 <MAN>  
A:Cross-references: GB:M76482; NID:G190751; PIDN:AAA0230.1; PID:G190752  
C:Genetics:  
A:Gene: GDB:IDSG3  
A:Cross-references: GDB:134030; OMTM:169615  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-999/Product: desmoglein homolog #status predicted <MAT>  
F:50-615/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-267/Domain: cadherin repeat homology <CR2>  
F:270-383/Domain: cadherin repeat homology <CR3>  
F:390-485/Domain: cadherin repeat homology <CR4>  
F:496-598/Domain: cadherin repeat homology <CR5>  
F:616-633/Domain: transmembrane #status predicted <TM>  
F:640-999/Domain: intracellular #status predicted <INT>  
F:910-938/Domain: desmoglein repeat <DG1>  
F:937-966/Domain: desmoglein repeat <DG2>  
F:110,180,343/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 82; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 2e+06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNRRYGPYTF 15  
DB 512 SARTLNRRYGPYTF 526

RESULT 2  
C70333  
hypothetical protein ag.372 - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 08-May-1998  
C:Accession: C70333  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
V.  
Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MUID:98196666  
A:Accession: C70333

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-531 <AOF>  
 A:Cross-references: GB:AE000686; NID:g2983038; PID:g2983043; GB:AE000657  
 A:Experimental source: strain VFS  
 C:Genetics:  
 A:Gene: aq\_372

Query Match 56.1%; Score 46; DB 2; Length 531;  
 Best Local Similarity 60.0%; Pred. No. 2.2;  
 Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 SARTLNRRYGPYTF 15  
 | | | | |  
 DB 77 SIETLNRRLEPYTF 91

RESULT 3  
 B70834  
 Hypothetical protein RV2991 - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: B70834  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A:Authors: Sares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
 A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MVID:98255987  
 A:Accession: B70834  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-163 <COL>  
 A:Cross-references: GB:AL021287; GB:AL123456; NID:g3261508; PID:el237736; PID:g2791588  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV2991

Query Match 51.2%; Score 42; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 RYTGPTT 14  
 | | | | |  
 DB 112 RYTGPTT 118

RESULT 4  
 A44282  
 retrovirus-related pol polyprotein pseudogene - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 31-Dec-1993 #sequence\_revision 04-Jan-1996 #text\_change 04-Jan-1996  
 C:Accession: A44282  
 R:Hirose, Y.; Takamatsu, M.; Harada, F.  
 Virology 192, 52-61, 1993  
 A>Title: Presence of env genes in members of the RTV-H family of human endogenous retro  
 A:Reference number: A44282; MVID:93297118  
 A:Accession: A44282  
 A:Status: conceptual translation of pseudogene  
 A:Molecule type: mRNA  
 A:Residues: 1-848 <HIR>  
 A:Cross-references: GB:D10083  
 C:Keywords: pseudogene  
 F:45/Region: opal stop codon  
 F:485/Region: opal stop codon  
 F:694/Region: ochre stop codon  
 F:815/Region: opal stop codon  
 F:827/Region: ochre stop codon  
 F:835/Region: opal stop codon

Query Match 48.8%; Score 40; DB 4; Length 848;  
 Best Local Similarity 63.6%; Pred. No. 4.2;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 3 RTLNRRYGPYTF 13  
 | | | | |  
 DB 660 QTLQPRWGPY 670

RESULT 5  
 A42548  
 genome polyprotein - mumps virus (strain Miyahara)  
 N:Alternate names: L protein  
 N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)  
 C:Species: mumps virus  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 11-Jun-1999  
 C:Accession: A42548  
 R:Okazaki, K.; Tanabayashi, K.; Takeuchi, K.; Hishiyama, M.; Okazaki, K.; Yamada, A.  
 Virology 188, 926-930, 1992  
 A>Title: Molecular cloning and sequence analysis of the mumps virus gene encoding the  
 A:Reference number: A42548; MVID:92263804  
 A:Accession: A42548  
 A:Molecule type: genomic RNA  
 A:Residues: 1-2261 <OKA>  
 A:Cross-references: GB:D10575; NID:g222145; PIDN:BA01432.1; PID:d1001906; PID:g22214  
 C:Genetics:  
 A:Gene: L  
 C:Superfamily: parainfluenza virus RNA-directed RNA polymerase  
 C:Keywords: ATP, nucleotidyltransferase

Query Match 47.6%; Score 39; DB 1; Length 2261;  
 Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 ARTLNRRYGPYTF 15  
 | | | | |  
 DB 686 ARTLNRRYGPYTF 699

RESULT 6  
 S16664  
 large protein L - parainfluenza virus type 2  
 C:Species: parainfluenza virus type 2  
 C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 18-Jun-1999  
 C:Accession: S16664  
 R:Kawano, M.; Okamoto, K.; Bando, H.; Kondo, K.; Tsurudome, M.; Komada, H.; Nishio, M.  
 Nucleic Acids Res. 19, 2739-2746, 1991  
 A>Title: Characterizations of the human parainfluenza type 2 virus gene encoding the  
 A:Reference number: S16659; MVID:91252221  
 A:Accession: S16664  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2262 <KAW>  
 A:Cross-references: EMBL:X57559; NID:961965; PIDN:CAA0788.1; PID:g61991  
 C:Superfamily: parainfluenza virus RNA-directed RNA polymerase

Query Match 47.6%; Score 39; DB 2; Length 2262;  
 Best Local Similarity 57.1%; Pred. No. 1.9e+02;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 ARTLNRRYGPYTF 15  
 | | | | |  
 DB 685 ARTLNRRYGPYTF 698

RESULT 7  
 S83577  
 C28A5.4 protein (clone C28A5) - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 07-Sep-1994 #sequence\_revision 10-Nov-1995 #text\_change 17-Oct-1997



A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-431 <KIS>  
 A:Cross-References: EMBL:U21729; NID:g717137; PID:g717138  
 C:Superfamily: glucagon receptor

Query Match 46.3%; Score 38; DB 2; Length 431;  
 Best Local Similarity 46.2%; Pred. No. 44;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 3 RTLNRYTGPYTF 15  
 |||:|:|:|:|:  
 DB 47 RTTGNFSGPYTY 59

RESULT 12  
 F71151  
 Probable glucose-1-phosphate thymidyltransferase - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 29-Sep-1999  
 C:Accession: F71151  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: F71151  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-356 <KAW>  
 A:Cross-References: GB:AP000002; NID:g3236129; PID:BA29499.1; PID:d1030442; PID:g323568  
 A:Experimental source: strain 073  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH0413  
 C:Superfamily: mannose-1-phosphate guanylyltransferase

Query Match 46.3%; Score 38; DB 2; Length 356;  
 Best Local Similarity 60.0%; Pred. No. 36;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 LNNRYTGPYTF 14  
 :|:|:|:|:|:|:  
 DB 284 LNNRYTGPYTF 293

RESULT 13  
 C70912  
 Hypothetical protein RV0043c - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: C70912  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: C70912  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-244 <COL>  
 A:Cross-References: GB:Z80775; GB:AL123456; NID:g3250715; PID:e268216; PID:g1568582  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: RV0043c

Query Match 46.3%; Score 38; DB 2; Length 244;  
 Best Local Similarity 60.0%; Pred. No. 23;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 RTLNRYTGP 12  
 |||:|:|:|:  
 DB 134 RTVNDYAGP 143

RESULT 14  
 B70931  
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 17-Jul-1998  
 C:Accession: B70931  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: B70931  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-423 <COL>  
 A:Cross-References: GB:AL020201; GB:AL123456; NID:g3250699; PID:e1254612; PID:g323448  
 A:Experimental source: strain H37RV  
 C:Genetics:  
 A:Gene: PPE

Query Match 46.3%; Score 38; DB 2; Length 423;  
 Best Local Similarity 63.6%; Pred. No. 43;  
 Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 LNNRYTGPYTF 15  
 |:|:|:|:|:|:  
 DB 241 LGGAVTGPYTF 251

RESULT 15  
 E69312  
 Nitrate reductase gamma subunit homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 05-Jun-1998  
 C:Accession: E69312  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod  
 F. Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirsch, E.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,  
 Smith, H.O.; Woese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: E69312  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-332 <KLE>  
 A:Cross-References: GB:AE001069; GB:AE000782; NID:g2689392; PID:g2650124; TIGR:AF0501

Query Match 46.3%; Score 38; DB 2; Length 332;  
 Best Local Similarity 46.2%; Pred. No. 33;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 3 RTLNRYTGPYTF 15  
 |||:|:|:|:|:  
 DB 53 RTYDRFSDPYTW 65

Search completed: January 7, 2000, 08:52:46  
 Job time: 2276 sec

Sat Jan 15 11:45:06 2000

us-08-991-628-6.rpr

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:40 : Search time 28.55 Seconds  
(without alignments)  
15.282 Million cell updates/sec

Title: US-08-991-628-6  
Perfect score: 82  
Sequence: 1 SARTLNNTGPTTF 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38:\*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	82	100.0	999	1	DSG3_HUMAN	P32926 homo sapien
2	41	50.0	391	1	GABR_XENLA	P00005 xenopus lae
3	39	47.6	245	1	GVPL_ANAFL	P55149 anabaena fl
4	39	47.6	273	1	HMD1_CAEEL	Q18273 caenorhabd
5	39	47.6	431	1	IM44_YEAST	Q01852 saccharomyc
6	39	47.6	2261	1	RRPL_XMMPM	P30929 numps vitus-
7	39	47.6	2262	1	RRPL_PI2HT	P26676 human paral
8	38	46.3	431	1	CRF2_MOUSE	Q60748 mus muscula
9	38	46.3	1111	1	SYLC_SCHPO	Q10490 schizosacch
10	38	46.3	244	1	YOH6_MYCTU	P71700 mycobacteri
11	37.5	45.7	406	1	YD11_SCHPO	P87049 schizosacch
12	37	45.1	387	1	ASPP_AEDAE	Q03168 aedes aegypt
13	37	45.1	413	1	ESCL_SCHPO	Q04635 schizosacch
14	37	45.1	643	1	SYRM_YEAST	P38714 saccharomyc
15	37	45.1	302	1	YDAK_ECOLI	P77744 escherichia
16	36	43.9	1417	1	BLM_HUMAN	P54132 homo sapien
17	36	43.9	504	1	C6B2_HELAM	Q27654 helicoverpa
18	36	43.9	362	1	COA1_POVBA	P14996 polyomaviru
19	36	43.9	340	1	COA1_POVMK	P24595 mouse polyo
20	36	43.9	372	1	COND_NEUCR	P19463 neuropepora
21	36	43.9	1337	1	POLA_DROME	Q10394 drosophila
22	36	43.9	3898	1	RRPL_HCVB	P21330 hog cholera
23	36	43.9	2269	1	RRPL_SV41	P35341 simian viru
24	36	43.9	334	1	VG08_BPTA	P19062 bacterioph
25	35	42.7	279	1	ASFL_YEAST	P32472 saccharomyc
26	35	42.7	383	1	COA1_POVM3	P03091 mouse polyo
27	35	42.7	382	1	COA1_POVMA	P03090 mouse polyo
28	35	42.7	383	1	COA1_POVMC	P12907 mouse polyo
29	35	42.7	383	1	COA1_POVMP	P49302 mouse polyo
30	35	42.7	325	1	KC1A_BOVIN	P35506 bos taurus
31	35	42.7	337	1	KC1A_CHICK	P70065 gallus gall
32	35	42.7	337	1	KC1A_HUMAN	P48729 homo sapien
33	35	42.7	325	1	KC1A_RAT	P97633 rattus norv
34	35	42.7	108	1	KVLE_HUMAN	P01597 homo sapien
35	35	42.7	469	1	NIFD_METTH	O27605 methanobact
36	35	42.7	4427	1	PKSL_BACSU	Q05470 bacillus su
37	35	42.7	120	1	POL1_MOUSE	P10400 mus musculu
38	35	42.7	390	1	POL3_MOUSE	P11367 mus musculu
39	35	42.7	843	1	POL_MLYAK	P03357 akr murine

## ALIGNMENTS

RESULT	1	STANDARD	PRT	999 AA
DSG3_HUMAN	1			
ID	DSG3_HUMAN			
AC	P32926			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 92069753.			
RA	AMAGAI M., KLAUS-KOVTUN V., STANLEY J.R.;			
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion."			
RL	Cell 67:869-877(1991).			
CC	- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.			
CC	- INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.			
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND CARCINOMAS.			
CC	- DOMAIN: CALCIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).			
CC	- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES AGAINST DSG3.			
CC	- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; M76482; AAA60230.1; -			
CC	PIR; A41088; IGH03.			
CC	HSSP; P09803; 1EDH.			
DR	MIM; 169615; -			
DR	PFAM; PF00028; cadherin; 4.			
DR	PROSITE; PS00232; CADHERIN; 3.			
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; Calcium-binding; Repeat.			
FT	SIGNAL	1	23	POTENTIAL.
FT	PROPEP	24	49	POTENTIAL.
FT	CHAIN	50	999	DESMOGLEIN 3.
FT	DOMAIN	50	615	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	616	640	POTENTIAL.
FT	DOMAIN	641	999	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	50	158	CADHERIN 1.
FT	REPEAT	159	268	CADHERIN 2.
FT	REPEAT	269	363	CADHERIN 3.
FT	REPEAT	364	499	CADHERIN 4.
FT	REPEAT	386	935	DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966	DESMOGLEIN REPEAT 2.
FT	REPEAT	967	110	POTENTIAL.
FT	CARBOHYD	110	110	

FT CARBOHYD 180 180 POTENTIAL.  
 FT CARBOHYD 459 459 POTENTIAL.  
 FT CARBOHYD 545 545 POTENTIAL.  
 SO SEQUENCE 999 AA; 107503 MW; 4891f6ae CRC32;

Query Match 100.0%; Score 82; DB 1; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 7, 6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SARTLNRYTGPYTF 15  
 |||||  
 DB 512 SARTLNRYTGPYTF 526

RESULT 2  
 ID GAB\_XENLA STANDARD; PRT; 391 AA.  
 AC P70005;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE TRANSCRIPTION FACTOR XGATA-6b (GATA BINDING FACTOR-6b).  
 GN GATA-6a OR GATA-6.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
 OC Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae; Xenopodinae;  
 OC Xenopus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER.  
 RX MEDLINE; 97180937.  
 RA GOVE C.D., WALMSLEY M., NIJAR S., BERTWISTLE D., GUILLE M.,  
 RA PATTINGTON G., BOMFORD A., PATIENT R.,  
 RT "Over-expression of GATA-6 in Xenopus embryos blocks differentiation  
 of heart precursors."  
 RL EMOB J. 16:355-368(1997).  
 CC -1- FUNCTION: ASSOCIATED WITH CARDIAC SPECIFICATION AND CAN REGULATE  
 CARDIAC-SPECIFIC TRANSCRIPTION DURING EMBRYOGENESIS. ACTIVATES  
 THE EXPRESSION OF CARDIAC MHC-ALPHA IN VIVO.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- SIMILARITY: BELONGS TO THE GATA-TYPE TRANSCRIPTION FACTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Y08865; CAA70088.1; .  
 CC DR HSSP; P17678; IGAU.  
 CC DR PFAM; PF00320; GATA. 2.  
 CC DR PROSITE; PS00344; GATA\_ZN\_FINGER; 2.  
 KM Transcription regulation; Activator; DNA-binding; Zinc-finger;  
 KM Nuclear protein.  
 FT ZN\_FING 182 206 GATA-TYPE.  
 FT ZN\_FING 236 260 GATA-TYPE.  
 FT DOMAIN 70 79 POLY-SER.  
 FT DOMAIN 241 245 POLY-THR.  
 FT DOMAIN 293 298 POLY-SER.  
 SO SEQUENCE 391 AA; 41503 MW; 221DDA28 CRC32;

Query Match 50.0%; Score 41; DB 1; Length 391;  
 Best Local Similarity 61.5%; Pred. No. 5, 5;  
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 ARTLNRYTGPYTF 14  
 |||||  
 DB 120 ARPLNGSYGSPYTF 132

RESULT 3  
 ID GVPL\_ANAFL STANDARD; PRT; 245 AA.  
 AC P55149;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE GVPE/L PROTEIN.  
 GN GVPE/L.  
 OS Anabaena flos-aquae.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CCAP 1403/13F;  
 RX MEDLINE; 97217685.  
 RA KINSMAN R., HAYES P.R.;  
 RT "Genes encoding proteins homologous to halobacterial Gyps N, J, K, F  
 & L are located downstream of gvpC in the cyanobacterium Anabaena  
 flos-aquae."  
 RL DNA Seq. 7:97-106(1997).  
 CC -1- FUNCTION: MAY PLAY A STRUCTURAL OR REGULATORY ROLE IN GAS  
 VESICLE SYNTHESIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; U17109; AAA58714.1; .  
 CC KM Gas vesicle.  
 SO SEQUENCE 245 AA; 28676 MW; F468D20E CRC32;

Query Match 47.6%; Score 39; DB 1; Length 245;  
 Best Local Similarity 66.7%; Pred. No. 7, 3;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 7 NRYTGPYTF 15  
 |||||  
 DB 232 NNFRAPYTF 240

RESULT 4  
 ID HMD1\_CAEL STANDARD; PRT; 273 AA.  
 AC Q18273;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PUTATIVE HOMEBOX PROTEIN C28A5.4.  
 GN C28A5.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA PALMER S.;  
 RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE DISTAL-LESS FAMILY OF HOMEBOX  
 PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



```

CC      or send an email to license@lsb-sib.ch).
CC      EMBL: 232680; CA83601.1;
CC      DR HSSP: P02836; 1HDD.
CC      DR WORMREP: C28A5.4; GE05325.
CC      DR PFAM: PF00046; homeobox.1.
CC      DR PROSITE: PS00027; HOMEBOX.1; 1.
CC      DR PROSITE: PS50071; HOMEBOX.2; 1.
CC      KW Hypothetical protein; Homeobox; DNA-binding; Nuclear protein.
CC      FT DNA_BIND 102 161 HOMEBOX.
CC      SQ SEQUENCE 273 AA; 30163 MW; 1665A6DF CRC32;

Query Match
Best Local Similarity 47.6%; Score 39; DB 1; Length 273;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      4 TLNNRYTGYPTF 15
DB      218 TLNEQYASPLYL 229

RESULT 5
ID      IM44_YEAST STANDARD; PRT; 431 AA.
AC      001852;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DE      15-JUL-1993 (Rel. 38, Last annotation update)
DE      MITOCHONDRIAL IMPORT INNER MEMBRANE TRANSLOCASE SUBUNIT TIM44
DE      PRECURSOR (MITOCHONDRIAL PROTEIN IMPORT PROTEIN 1) (INNER MEMBRANE
DE      IMPORT SITE PROTEIN 45) (ISP45) (MEMBRANE IMPORT MACHINERY PROTEIN
DE      MIM44)
GN      TIM44 OR MPT1 OR MIM44 OR ISP45 OR YII022W.
OS      Saccharomycetes cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC      Saccharomycetaceae; Saccharomyces.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AB320;
RX      MEDLINE: 931010953.
RA      MAARSE A.C., BLOW J., GRIVELL L.A., MEIJER M.;
RT      "Mpt1, an essential gene encoding a mitochondrial membrane protein,
RT      is possibly involved in protein import into yeast mitochondria.",
RL      EMBL J. 113619-3628(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN-S288C / AB972;
RA      BARRELL B.G., BADOOCK K., BANKIER A.T., BOWMAN S., BROWN D.,
RA      CHURCHER C.M., CONNOR R., CORSEY T., DEAR S., DEVLIN K., FRASER A.,
RA      GENTLES S., HAMLIN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,
RA      LOUIS E., LYE G., MOOLE S., MOULE T., ODELL C., PEARSON D.,
RA      RAVANDEHAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,
RA      WALSH S.V., WHITEHEAD S.;
RL      Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RN      [3]
RP      PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX      MEDLINE: 93345448.
RA      HORST M., JENOE P., KRONIDOU N.G., BOLLIGER L., OPLIGER W.,
RA      SCHERER P., MANNING-KRIEG U., JASCUR T., SCHATZ G.;
RT      "Protein import into yeast mitochondria: the inner membrane import
RT      site protein ISP45 is the MPT1 gene product.",
RL      EMBL J. 123035-3041(1993).
CC      -1- FUNCTION: INVOLVED IN PROTEIN IMPORT INTO THE MITOCHONDRION.
CC      PROBABLY INVOLVED IN TRANSLLOCATION DURING THE IMPORT OF
CC      AS A BINDING PROTEIN REQUIRED FOR DRIVING THE IMPORT OF
CC      PREPROTEINS. RECRUITS MITOCHONDRIAL HSP70 AND ITS CO-CHAPERONE
CC      (HSP71) TO DRIVE PROTEIN TRANSLLOCATION INTO THE MATRIX USING ATP
CC      AS AN ENERGY SOURCE.
CC      -1- SUBUNIT: FORMS PART OF THE RECEPTOR COMPLEX THAT CONSISTS OF AT
CC      LEAST 3 DIFFERENT PROTEINS (TIM17, TIM23, TIM44).
CC      -1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC      -1- SIMILARITY: BELONGS TO THE TIM44 FAMILY.

```

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
CC      EMBL: 246881; CA86970.1;
CC      DR EMBL: X67276; CA87693.1;
CC      DR PIR: S25196; S25196.
CC      DR SGD: L0001138; MPT1.
CC      KW Mitochondrion; Inner Membrane; Transport; Protein transport;
CC      Translocation; Transic peptide; ATP-binding.
CC      FT TRANSIT 1 431 MITOCHONDRIAL IMPORT INNER MEMBRANE
CC      CHAIN ? 431 TRANSLOCASE SUBUNIT TIM44.
CC      FT NP_BIND 101 108 ATP (POTENTIAL).
CC      SQ SEQUENCE 431 AA; 48854 MW; 002E0771 CRC32;

Query Match
Best Local Similarity 47.6%; Score 39; DB 1; Length 431;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      1 SARTLNRYTGYPT 14
DB      13 SSRITLRYRQYPT 26

RESULT 6
ID      RPL_XMMPM STANDARD; PRT; 2261 AA.
AC      P30929;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-JUL-1993 (Rel. 26, Last sequence update)
DE      15-DEC-1998 (Rel. 37, Last annotation update)
DE      RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)
DE      (L PROTEIN).
GN      L.
OS      Mumps virus (strain Miyahara vaccine).
OC      Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC      Paramyxoviridae; Paramyxovirinae; Rubulavirus.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE: 92263804.
RA      OKAZAKI K., TANABAYASHI K., TAKEUCHI K., HISHIYAMA M., OKAZAKI K.;
RA      YAMADA A.;
RT      "Molecular cloning and sequence analysis of the mumps virus gene
RT      encoding the L protein and the trailer sequence.",
RL      Virology 188:926-930(1992).
CC      -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY
CC      SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs. RNA EDITING OF THE P
CC      GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.
CC      -1- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@lsb-sib.ch).
CC      EMBL: D10575; BA01432.1;
CC      DR PIR: A42548; A42548.
CC      DR PFAM: PF00946; Paramyx-RNA-pol.1.
CC      KW Transferase; RNT-directed RNA polymerase.
CC      SQ SEQUENCE 2261 AA; 256572 MW; A6C5C12 CRC32;

```

Query Match 47.6%; Score 39; DB 1; Length 2261;  
 Best Local Similarity 57.1%; Pred. No. 91;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 ARTLNRYTGPYTF 15  
 686 ARTLNRYTGPYTF 699

RESULT 7  
 RRP1\_P12HT STANDARD; PRT: 2262 AA.  
 AC P2676;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE RNA POLYMERASE BETA SUBUNIT (EC 2.7.7.48) (LARGE STRUCTURAL PROTEIN)  
 DE (L PROTEIN).  
 GN L.  
 OS Human parainfluenza 2 virus (strain Toshida) (PIV-2).  
 OC Viruses: ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Rubulavirinae;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91252221.  
 RA KAWANO M., OKAMOTO K., BANDO H., KONDO K., TSURUDOME M., KOMADA H.,  
 RA NISHIO M., ITO Y.;  
 RT "Characterizations of the human parainfluenza type 2 virus gene  
 RT encoding the L protein and the intergenic sequences.";  
 RL Nucleic Acids Res. 19:2739-2746(1991).  
 CC -1- FUNCTION: PROBABLE COMPONENT OF THE ACTIVE POLYMERASE. IT MAY  
 CC FUNCTION IN MRNA SYNTHESIS, CAPPING, METHYLATION AND POLY(A)  
 CC SYNTHESIS OF NEWLY SYNTHESIZED VIRAL MRNAs. RNA EDITING OF THE P  
 CC GENE TRANSCRIPT, AND PROTEIN KINASE ACTIVITY.  
 CC -1- SIMILARITY: WITH L PROTEIN OF OTHER PARAMYXOVIRUSES.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X57559; CAA40788.1; -  
 DR PIR: S16664; S16664  
 DR PIR: PF00946; Paramyx\_RNA\_pol.1.  
 DR Transferase; RNA-directed RNA polymerase.  
 KW SEQUENCE 2262 AA; 256380 MW; 2EF04669 CRC32;

Query Match 47.6%; Score 39; DB 1; Length 2262;  
 Best Local Similarity 57.1%; Pred. No. 91;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 2 ARTLNRYTGPYTF 15  
 685 ARTLNRYTGPYTF 698

RESULT 8  
 CRF2\_MOUSE STANDARD; PRT: 431 AA.  
 AC 060748; 060808; 060783;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE CORTICOTROPIN-RELEASING FACTOR RECEPTOR 2 PRECURSOR (CRF-R) (CRF2)  
 DE (CRF-R) (CRF-R2).  
 GN CRHR2 OR CRF2R.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-HEART;  
 RX MEDLINE: 95224061.  
 RA PERRIN M., DONALDSON C., CHEN R., BLOUNT A., BERGGREN T.,  
 RA BILIZIKIAN L., SAMCHENKO P., VALE W.;  
 RT "Identification of a second corticotropin-releasing factor receptor  
 RT gene and characterization of a cDNA expressed in heart.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:2969-2973(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-HEART;  
 RX MEDLINE: 95166778.  
 RA KISHIMOTO T., PEARSE R.V. II, LIN C.R., ROSENFIELD M.G.;  
 RT "A sauvagie/corticotropin-releasing factor receptor expressed in  
 RT heart and skeletal muscle.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:1108-1112(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-HEART;  
 RX MEDLINE: 96015396.  
 RA STENZEL P., KESTERSON R., YEUNG W., CONE R.D., RITTENBERG M.B.,  
 RA STENZEL-POORE M.P.;  
 RT "Identification of a novel murine receptor for  
 RT corticotropin-releasing hormone expressed in the heart.";  
 RL Mol. Endocrinol. 9:637-645(1995).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR CORTICOTROPIN-RELEASING FACTOR.  
 CC SHOWS HIGH-AFFINITY CRF BINDING. THE ACTIVITY OF THIS RECEPTOR IS  
 CC MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL CYCLASE.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE HEART. ALSO EXPRESSED  
 CC IN LUNGS, SKELETAL MUSCLE, GASTROINTESTINAL TRACT, EPIDIDYMIS, AND  
 CC BRAIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U17858; AAA68026.1; -  
 DR EMBL: U21729; AAC52174.1; -  
 DR EMBL: U19939; AAC52243.1; -  
 DR GCRDB: GCR\_1705; -  
 DR GCRDB: GCR\_1711; -  
 DR GCRDB: GCR\_184312; CRHR2.  
 DR PIR: PF00002; 7tm\_2; 1.  
 DR PROSITE: PS00649; G-PROTEIN-RECEP\_F2\_1; 1.  
 DR PROSITE: PS00650; G-PROTEIN-RECEP\_F2\_2; 1.  
 KW G-protein coupled receptor; Glycoprotein; Signal.  
 FT SIGNAL 1 24  
 FT CHAIN 25 431  
 FT DOMAIN 25 138  
 FT TRANSSEM 139 159  
 FT DOMAIN 160 168  
 FT TRANSSEM 169 188  
 FT DOMAIN 189 205  
 FT TRANSSEM 206 229  
 FT DOMAIN 230 243  
 FT TRANSSEM 244 265  
 FT DOMAIN 266 284  
 FT TRANSSEM 285 307  
 FT DOMAIN 308 330  
 FT TRANSSEM 331 350  
 FT DOMAIN 351 365  
 FT TRANSSEM 366 385  
 FT DOMAIN 386 431  
 FT CARBOHYD 52 52  
 FT CARBOHYD 61 61

EXTRACELLULAR (POTENTIAL).  
 1 (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 2 (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 3 (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 4 (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 5 (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 6 (POTENTIAL).  
 EXTRACELLULAR (POTENTIAL).  
 7 (POTENTIAL).  
 CYTOPLASMIC (POTENTIAL).  
 POTENTIAL.

OY 1 OSGTMRHSTGCTN 15  
 DB 387 ODCALNTGHTSGVN 401

## RESULT 4

HEMA\_P13B STANDARD: PRT: 572 AA.

AC P06167:

DT 01-JAN-1988 (Rel. 06, Last sequence update)

DT 01-OCT-1996 (Rel. 34, Last annotation update)

DE HEMAGGLUTININ-NEURAMINIDASE (EC 3.2.1.18).

GN HN.

OS Bovine parainfluenza 3 virus.

OC Viruses; ssRNA negative-strand viruses; Mononegavirales;

OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-910N:

RA MEDLINE: 87174819.

RT SUZU S., SAKAI Y., SHIODA T., SHIBUTA H.;

RT "Nucleotide sequence of the bovine parainfluenza 3 virus genome: the

RT genes of the F and HN glycoproteins."

RL Nucleic Acids Res. 15:2945-2958(1987).

CC -1- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS

CC TO CELL RECEPTORS AND FOR INITIATING INFECTION. NEURAMINIDASE

CC ACTIVITY HELPS THE EFFICIENT SPREAD OF THE VIRUS BY DISSOCIATING

CC THE MATURE VIRIONS FROM THE NEURAMINIC ACID CONTAINING

CC GLYCOPROTEINS.

CC -1- SUBCELLULAR LOCATION: EXTERNAL, ANCHORED TO THE ENVELOPE BY ITS

CC -1- SIMILARITY: BELONGS TO THE PARAMYXOVIRUSES HEMAGGLUTININ-

CC NEURAMINIDASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: Y00114; CAA68298.1;

DR EMBL: D84095; BAA12218.1;

DR PIR: B37218; HNNZB3.

DR PIR: PF00423; HN7.1.

KM Hydrolase; Hemagglutinin; Envelope protein; Glycoprotein;

KM Transmembrane.

FT DOMAIN 1 30

FT TRANSMEM 31 53

FT DOMAIN 54 572

FT CARBOHYD 308 308

FT CARBOHYD 351 351

FT CARBOHYD 448 448

FT CARBOHYD 523 523

FT CARBOHYD 570 570

FT CARBOHYD 570 570

SEQUENCE 572 AA: 64590 MW; 876CFE66 CRC32;

OY 3 GIMRHRHSTGCTN 15

DB 18 GTRDRHSSKATN 30

RESULT 5

AMY\_BACLI

ID AMY\_BACLI

STANDARD: PRT: 512 AA.

Query Match 51.2%; Score 41; DB 1; Length 572;

Best Local Similarity 61.5%; Pred. No. 7.3;

Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

P06278:  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ALPHA-AMYLASE PRECURSOR (EC 3.2.1.1) (1.4-ALPHA-D-GLUCAN  
 DE GLUCANHYDROLASE).  
 GN AMY OR AMYL.  
 OS Bacillus licheniformis.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 27811;  
 RX MEDLINE: 86111694.  
 RA YUKI T., NOMURA T., TEZUKA H., TSUBOI A., YAMAGATA H.,  
 RA TSUKAGOSHI N., UDAKA S.;  
 RT "Complete nucleotide sequence of a gene coding for heat- and  
 RT pH-stable alpha-amylase of Bacillus licheniformis: comparison of the  
 RT amino acid sequences of three bacterial liquefying alpha-amylases  
 RT deduced from the DNA sequences."  
 RL J. Biochem. 98:1147-1156(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86195857.  
 RA GRAY G.L., MAINZER S.E., REY M.W., LAMSA M.H., KINDLE K.L.,  
 RA CARONA C., ROUDADT C.;  
 RT "Structural genes encoding the thermophilic alpha-amylases of  
 RT Bacillus stearothermophilus and Bacillus licheniformis."  
 RL J. Bacteriol. 166:635-643(1986).  
 RN [3]  
 RP SEQUENCE OF 1-104 FROM N.A.  
 RX MEDLINE: 84185455.  
 RA STEPHENS M.A., ORTLEPP S.A., OLLINGTON J.F., MCCONNELL D.J.;  
 RT "Nucleotide sequence of the 5' region of the Bacillus licheniformis  
 RT alpha-amylase gene: comparison with the B. amyloliquefaciens gene."  
 RL J. Bacteriol. 158:369-372(1984).  
 RN [4]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE: 89213924.  
 RA LAODE B.M., CHAVELISS G.H., MCCONNELL D.J.;  
 RT "Bacillus licheniformis alpha-amylase gene, amyl, is subject to  
 RT promoter-independent catabolite repression in Bacillus subtilis."  
 RL J. Bacteriol. 171:2435-2442(1989).  
 RN [5]  
 RP SEQUENCE OF 30-47.  
 RX MEDLINE: 82098050.  
 RA KUHN H., FRETZER P.P., LAMPEN J.O.;  
 RT "N-terminal amino acid sequence of Bacillus licheniformis  
 RT alpha-amylase: comparison with Bacillus amyloliquefaciens and  
 RT Bacillus subtilis Enzymes."  
 RL J. Bacteriol. 149:372-373(1982).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).  
 RC STRAIN-ATCC 27811;  
 RX MEDLINE: 95182462.  
 RA MACHARD M., WIEGAND G., HUBER R.;  
 RT "Crystal structure of calcium-depleted Bacillus licheniformis alpha-  
 RT amylase at 2.2-A resolution."  
 RL J. Mol. Biol. 246:545-559(1995).  
 CC -1- CATALYTIC ACTIVITY: ENOXYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC  
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO  
 CC KNOWN AS THE ALPHA-AMYLASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: X03236; CAA26981.1;

DR EMBL: M38570; AAA22226.1; -  
 DR EMBL: M13256; AAA22240.1; -  
 DR EMBL: K01984; AAA22193.1; -  
 DR EMBL: M26412; AAA22237.1; -  
 DR EMBL: A17930; CAA01355.1; -  
 DR PIR: A00844; ALBSL.  
 DR PIR: B24549; B24549.  
 DR PIR: A26151; A26151.  
 DR PDB: 1BPL; 17-AUG-96.  
 DR PDB: 1VIS; 12-MAR-97.  
 DR PFAM: PF00128; alpha-amyrase; 1.  
 DR Hydrolase; Glycosidase; Carbohydrate metabolism; signal; 3D-structure.  
 FT SIGNAL 1 29  
 FT CHAIN 30 512 ALPHA-AMYLASE.  
 FT ACT\_SITE 260 260 BY SIMILARITY.  
 FT ACT\_SITE 264 264 BY SIMILARITY.  
 FT ACT\_SITE 357 357 BY SIMILARITY.  
 FT CONFLICT 38 38 Q -> Y (IN REF. 5).  
 FT CONFLICT 163 163 R -> L (IN REF. 2).  
 FT CONFLICT 339 339 S -> G (IN REF. 2).  
 FT CONFLICT 349 349 A -> S (IN REF. 2).  
 SQ SEQUENCE 512 AA; 58549 MW; 4B5D334D CRC32;

Query Match 50.0%; Score 40; DB 1; Length 512;  
 Best Local Similarity 53.8%; Pred. No. 9.7;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 QSGTMRTRHSTG 13  
 Db 98 QKGTVRKTKG 110

RESULT 6  
 CATB\_PSEAE STANDARD; PRT; 513 AA.  
 ID CATB\_PSEAE  
 AC 059635;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE CATALASE PRECURSOR (EC 1.11.1.6) (PARADOX INDUCIBLE CATALASE ISOZYME B).  
 GN KATB.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;  
 RN Pseudomonas.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRD2;  
 RX MEDLINE: 96062238.  
 RA BROWN S.M., HOWELL M.L., VASIL M.L., ANDERSON A.J., HASSETT D.J.;  
 RT "Cloning and characterization of the katB gene of Pseudomonas  
 aeruginosa encoding a hydrogen peroxide-inducible catalase:  
 purification of katB, cellular localization, and demonstration that  
 it is essential for optimal resistance to hydrogen peroxide.";  
 RT J. Bacteriol. 177:6536-6544(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-FRD1;  
 RA HOWELL M.L., HEUR M., KLOTZ M.G., HASSETT D.J.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: DECOMPOSES HYDROGEN PEROXIDE IN WATER AND OXYGEN; SERVES  
 TO PROTECT CELLS FROM THE TOXIC EFFECTS OF HYDROGEN PEROXIDE.  
 CC -1- CATALYTIC ACTIVITY: 2 H(2)O(2) -> O(2) + 2 H(2)O.  
 CC -1- CORRECTOR: HEME GROUP.  
 CC -1- SIMILARITY: BELONGS TO THE CATALASE FAMILY.  
 CC -1- SUBCELLULAR LOCATION: PERIPLASMIC (PROBABLY).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: U34896; AAA79046.1; -  
 DR EMBL: U89384; AAB49463.1; -  
 DR HSPD: P00432; TCAT.  
 DR PFAM: PF00199; catalase; 1.  
 DR PROSITE: PS00437; CATALASE\_1; 1.  
 DR PROSITE: PS00438; CATALASE\_2; 1.  
 DR Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide;  
 KW periplasmic; signal.  
 FT SIGNAL 1 30  
 FT CHAIN 31 513 POTENTIAL.  
 FT ACT\_SITE 81 81 CATALASE.  
 FT ACT\_SITE 153 153 BY SIMILARITY.  
 FT BINDING 361 361 PROXIMAL HEME LIGAND (BY SIMILARITY).  
 SQ SEQUENCE 513 AA; 57131 MW; D18F8D CRC32;

Query Match 50.0%; Score 40; DB 1; Length 513;  
 Best Local Similarity 46.7%; Pred. No. 9.7;  
 Matches 7; Conservative 1; Mismatches 7; Indels 0; Gaps 0;  
 Oy 1 QSGTMRTRHSTG 15  
 Db 390 QDGALNAGHSTG 404

RESULT 7  
 FILA\_HUMAN STANDARD; PRT; 416 AA.  
 ID FILA\_HUMAN  
 AC P20930;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE FLAGGRIN PRECURSOR (FRAGMENT).  
 GN FLAG.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89296901.  
 RA MCKINLEY-GRANT L.J., IDLER W.W., BERNSTEIN I.A., PARRY D.A.D.,  
 RA CANNIZZARO L., CROCE C.M., HUBNER K., LESSIN S.R., STEINERT P.M.;  
 RT "Characterization of a cDNA clone encoding human flaggrin and  
 localization of the gene to chromosome region 1q21.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).  
 CC -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES  
 DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING  
 TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.  
 CC -1- PIN: FLAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,  
 HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES  
 OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE  
 PRECURSOR IS DEPOSITED AS KERATOHALIN GRANULES. DURING TERMINAL  
 DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTOLYTICALLY CLEAVED.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: M24355; AAA52454.1; -  
 DR PIR: A32947; A32947.  
 DR MIM: J35940; -  
 DR phosphorylation; Polyprotein; Developmental protein.  
 FT NON\_TER 1  
 SQ SEQUENCE 416 AA; 44105 MW; 72D28913 CRC32;

Query Match 50.0%; Score 40; DB 1; Length 416;  
 Best Local Similarity 61.5%; Pred. No. 7.8;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 QSGTMRHSTG 13  
 1111 1:1:1  
 DB 237 QSGTMRHSTG 249

RESULT 8  
 FYN\_CHICK STANDARD: PRT: 533 AA.  
 ID FYN\_CHICK  
 AC 005876; 1-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN)  
 GN FYN  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WHITE LEGHORN; TISSUE=MUSCLE;  
 RA MEDLINE: 93205395.  
 RA SUDOL M., GREULICH H., NEWMAN L., SARKAR A., SUKAGAWA J., YAMAMOTO T.;  
 RT "A novel yes-related kinase, yrk, is expressed at elevated levels in  
 RT neural and hematopoietic tissues.";  
 RL Oncogene 8:823-831(1993).  
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- TISSUE SPECIFICITY: THYMUS AND SPLEEN.  
 CC -1- SIMILARITY: CONTRAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTRAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X52841; CA937025.1;  
 DR PIR: S36351; S36351.  
 DR HSSP: P06241; 1FYN.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; PKINASE; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS50001; SH2; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW Proto-oncogene: Tyrosinase; Tyrosine-protein kinase; Phosphorylation;  
 KW ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;  
 KW Lipoprotein.  
 KW INIT\_MER 0 0  
 FT LIPID 1 1 MYRISTATE (BY SIMILARITY).  
 FT LIPID 2 2 PALMITATE (BY SIMILARITY).  
 FT LIPID 5 5 PALMITATE (BY SIMILARITY).  
 FT DOMAIN 81 142 SH3.  
 FT DOMAIN 148 245 SH2.  
 FT DOMAIN 267 520 PROTEIN KINASE.  
 FT MOD\_RES 11 11 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 FT NP\_BIND 273 281 ATP (BY SIMILARITY).  
 FT BINDING 295 295 ATP (BY SIMILARITY).  
 FT ACT\_SITE 386 386 BY SIMILARITY.

FT MOD\_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 527 527 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 533 AA; 60134 MM; 05970E9A CRC32;

Query Match 50.0%; Score 40; DB 1; Length 533;  
 Best Local Similarity 61.5%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SGTMRHSTG 14  
 1111 1:1:1  
 DB 72 TGTLRGTGT 84

RESULT 9  
 FYN\_HUMAN STANDARD: PRT: 536 AA.  
 ID FYN\_HUMAN  
 AC P06241; 1-JAN-1988 (Rel. 06, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1998 (Rel. 38, Last annotation update)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN)  
 GN FYN  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SEMBA K., NISHIZAWA M., MIYAJIMA N., YOSHIDA M.C., SUKAGAWA J.,  
 RA YAMAMOTO T., YAMAMOTO T., TOTOSHIMA K.;  
 RT "Yes-related proto-oncogene, syn, belongs to the protein-tyrosine  
 RT kinase family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:5459-5463(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 87089775.  
 RA KAWAKAMI T., PENNINGTON C.Y., ROBBINS K.C.;  
 RT "Isolation and oncogenic potential of a novel human src-like gene.";  
 RL Mol. Cell. Biol. 6:4195-4201(1986).  
 RN [3]  
 RP MYRISTOYLATION AND PHOSPHORYLATION AT TYR-530.  
 RC MEDLINE: 91016431.  
 RA PETERS D.J., MCGHEE B.R., PERON D.C., LIPTAK L.M., LAUDANO A.P.;  
 RT "In vivo phosphorylation and membrane association of the fyn proto-  
 RT oncogene product in IM-9 human lymphoblasts.";  
 RL Oncogene 5:1313-1319(1990).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF SH3 DOMAIN.  
 RC MEDLINE: 93327730.  
 RA NOBLE M.E.M., MUSACCHIO A., SARASTE M., COURTNEIDGE S.A.,  
 RA WIERENGA R.K.;  
 RT "Crystal structure of the SH3 domain in human Fyn; comparison of the  
 RT three-dimensional structures of SH3 domains in tyrosine kinases and  
 RT spectrin.";  
 RL EMBO J. 12:2617-2624(1993).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 80-141.  
 RC MEDLINE: 95393198.  
 RA MUSACCHIO A., SARASTE M., WILMANS M.;  
 RT "High-resolution crystal structures of tyrosine kinase SH3 domains  
 RT complexed with proline-rich peptides.";  
 RL Nat. Struct. Biol. 1:546-551(1994).  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 84-140 IN COMPLEX WITH NEF.  
 RC MEDLINE: 96279837.  
 RA LEE C.H., SAKSELA K., MIRZA U.A., CHAIT B.T., KURIYAN J.;  
 RT "Crystal structure of the conserved core of HIV-1 Nef complexed with  
 RT a Src family SH3 domain.";  
 RL Cell 85:931-942(1996).  
 RN [7]  
 RP STRUCTURE BY NMR OF SH3 DOMAIN.

RX MEDLINE: 96399716.  
 RA MORTON C.J., PUGH D.J.R., BROWN E.L.J., KAHMANN J.D., RENZONI D.A.C.,  
 RA CAMPBELL I.D.;  
 RT "solution structure and peptide binding of the SH3 domain from human  
 RT Fyn";  
 RL Structure 4:705-714(1996).  
 RN [8]  
 RP STRUCTURE BY NMR.  
 RX MEDLINE: 97121261.  
 RA RENZONI D.A., PUGH D.J., SILIGARDI G., DAS P., MORTON C.J., ROSSI C.,  
 RA WATERFIELD M.D., CAMPBELL I.D., LABURRY J.E.;  
 RT "structural and thermodynamic characterization of the interaction of  
 RT the SH3 domain from Fyn with the proline-rich binding site on the p85  
 RT subunit of PI3-kinase";  
 RL Biochemistry 35:15646-15653(1996).  
 RN [9]  
 RP STRUCTURE BY NMR OF SH2 DOMAIN.  
 RX MEDLINE: 98035454.  
 RA MULHERN T.D., SHAW G.L., MORTON C.J., DAY A.J., CAMPBELL I.D.;  
 RT "The SH2 domain from the tyrosine kinase Fyn in complex with a  
 RT phosphotyrosyl peptide reveals insights into domain stability and  
 RT binding specificity";  
 RL Structure 5:1313-1323(1997).  
 RN [10]  
 RP BINDING OF SH3 DOMAIN TO PI 3-KINASE.  
 RX MEDLINE: 93348274.  
 RA PRASAD K.V., JANSSEN O., KAPPELLER R., RAAB M., CANTLEY L.C.,  
 RA RUDD C.E.;  
 RT "Src-homology 3 domain of protein kinase p59fyn mediates binding to  
 RT phosphatidylinositol 3-kinase in T cells";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:7366-7370(1993).  
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sb.ch](mailto:license@sib-sb.ch)).  
 CC -----  
 DR EMBL: M14333; AAC08285.1; -  
 DR EMBL: M14676; AAA3615.1; -  
 DR PIR: A24314; TVHUSY.  
 DR PIR: A23389; TVHUSR.  
 DR PDB: 1SHF; 31-OCT-93.  
 DR PDB: 1FYN; 08-NOV-96.  
 DR PDB: 1NFF; 08-NOV-96.  
 DR PDB: 1NFG; 08-NOV-96.  
 DR PDB: 1FEN; 11-JAN-97.  
 DR PDB: 1AON; 25-FEB-98.  
 DR PDB: 1AOT; 14-JAN-98.  
 DR PDB: 1AOU; 14-JAN-98.  
 DR PDB: 1AGG; 25-FEB-98.  
 DR MIM; 137025; -  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; PKINASE; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS50011; SH2; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KM Proto-oncogene; Transferrase; Tyrosine-protein kinase; Phosphorylation;  
 KM ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;

KW Lipoprotein; 3D-structure.  
 FT INIT\_MET 0 0  
 FT LIPID 1 1  
 FT LIPID 2 2  
 FT LIPID 5 5  
 FT LIPID 142 142  
 FT DOMAIN 81 81  
 FT DOMAIN 148 148  
 FT DOMAIN 245 245  
 FT DOMAIN 270 270  
 FT MOD\_RES 11 11  
 FT NP\_BIND 276 284  
 FT BINDING 298 298  
 FT ACT\_SITE 389 389  
 FT MOD\_RES 419 419  
 FT MOD\_RES 530 530  
 FT MOD\_RES 530 530  
 FT CONFLICT 183 183  
 FT CONFLICT 436 436  
 FT STRAND 85 88  
 FT STRAND 92 92  
 FT STRAND 99 99  
 FT STRAND 102 102  
 FT STRAND 104 105  
 FT STRAND 107 112  
 FT STRAND 118 123  
 FT STRAND 124 126  
 FT STRAND 129 133  
 FT STRAND 134 136  
 FT HELIX 137 139  
 FT SEQUENCE 536 AA; 60630 MW; 01B18DD0 CRC32;  
 SQ  
 Query Match 50.0%; Score 40; DB 1; Length 536;  
 Best Local Similarity 61.5%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 SGMTRHSTGCT 14  
 DB 72 TGLTRGTGTVT 84  
 RESULT 10  
 FYN\_MOUSE STANDARD; PRT; 533 AA.  
 ID P39688; 01-FEB-1995 (Rel. 31, Created)  
 AC P39688; 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).  
 GN FYN  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91175680.  
 RA COOKE M.P., PERLMUTTER R.M.;  
 RT "Expression of a novel form of the fyn proto-oncogene in  
 RT hematopoietic cells";  
 RL New Biol. 1:66-74(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA LEE C., KIM M.G., JEON S.H., PARK D.E., PARK S.D., SEONG R.H.;  
 RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP PALMITOYLATION.  
 RX MEDLINE: 94019312.  
 RA SHEROV-SCARIA A.M., TIMSON L.K., KWONG J., SHAW A.S., LUBLIN D.M.;  
 RT "Palmitylation of an amino-terminal cysteine motif of protein  
 RT tyrosine kinases p56lck and p59fyn mediates interaction with  
 RT glycosyl-phosphatidylinositol-anchored proteins";  
 RL Mol. Cell. Biol. 13:6385-6392(1993).  
 RN [4]  
 RP PALMITOYLATION.  
 RX MEDLINE: 95071286.

RA KOEGL M., ZLATKINE P., LEY S.C., COURTNEIDGE S.A., MAGEE A.I.;  
 RT "Palmitoylation of multiple Src-family kinases at a homologous N-  
 terminal motif.";  
 RL Biochem. J. 303:749-753(1994).  
 (5)  
 RN RP PALMITOYLATION.  
 RX MEDLINE: 97345356.  
 RA MOYVEN A., OKAMURA H., ROSENBLATT Y., RESH M.D.;  
 RT "Palmitoylation of p59fyn is reversible and sufficient for plasma  
 membrane association.";  
 RL Mol. Biol. Cell 8:1159-1173(1997).  
 (6)  
 RN RP MYRISTOYLATION.  
 RX MEDLINE: 96251668.  
 RA GAUVEN L.K.T., LINDER M.E., SHAW A.S.;  
 RT "Multiple features of the p59fyn src homology 4 domain define a motif  
 for immune-receptor tyrosine-based activation motif (ITAM) binding  
 and for plasma membrane localization.";  
 RL J. Cell Biol. 133:1007-1015(1996).  
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M27266; AAA37644.1; -;  
 DR EMBL: U70324; AAB09568.1; -;  
 DR PIR: A44991; A44991.  
 DR HSSP: P06241; 1FVN.  
 DR MGD: MGI:95602; FVN.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; PKINASE; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS50001; SH2; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 DR PROTO-ONCOGENE; TRANSFERASE; TYROSINE-protein kinase; Phosphorylation;  
 KW ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;  
 KM Lipoprotein.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT LIPID 1 1 MYRISTATE.  
 FT LIPID 2 2 PALMITATE.  
 FT LIPID 5 5 PALMITATE.  
 FT LIPID 81 142 SH3.  
 FT DOMAIN 148 245 SH2.  
 FT DOMAIN 267 520 SH3.  
 FT DOMAIN 273 281 PROTEIN KINASE.  
 FT NP\_BIND 273 281 ATP (BY SIMILARITY).  
 FT BINDING 295 295 ATP (BY SIMILARITY).  
 FT ACT\_SITE 386 386 BY SIMILARITY.  
 FT MOD\_RES 416 416 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 533 AA; 59926 MW; B235B6FA CRC32;

Query Match 50.0%; Score 40; DB 1; Length 533;  
 Best Local Similarity 61.5%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SGTMRTHSGT 14  
 :||:|||||

DB 72 TGLTRRGITGV 84

RESULT 11

ID FYN\_XENLA STANDARD; PRT; 536 AA.

AC P13406;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).

GN FYN.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;

OC Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;

OC Xenopus.

RN (1)

RP SEQUENCE FROM N.A.

RX MEDLINE: 90191723.

RA STEELE R.E., DENG J.C., GHOSH C.R., FERRO J.B.;

RT "Structure and expression of fyn genes in Xenopus laevis.";

RL Oncogene 5:369-376(1990).

CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.

CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +

CC PROTEIN TYROSINE PHOSPHATE.

CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF

CC PHOSPHATIDYLINOSITOL 3-KINASE.

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC

CC DOMAIN. BELONGS TO THE SRC SUBFAMILY.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X52188; CAA36435.1; -;

DR EMBL: M27502; AAA49719.1; -;

DR PIR: A43806; A43806.

DR HSSP: P06241; 1FVN.

DR PFAM: PF00017; SH2; 1.

DR PFAM: PF00018; SH3; 1.

DR PFAM: PF00069; PKINASE; 1.

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE: PS50001; SH2; 1.

DR PROSITE: PS50002; SH3; 1.

DR PROTO-ONCOGENE; TRANSFERASE; TYROSINE-protein kinase; Phosphorylation;  
 KW ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;  
 KM Lipoprotein.  
 FT INIT\_MET 0 0 BY SIMILARITY.  
 FT LIPID 1 1 MYRISTATE (BY SIMILARITY).  
 FT LIPID 2 2 PALMITATE (BY SIMILARITY).  
 FT LIPID 5 5 PALMITATE (BY SIMILARITY).  
 FT LIPID 81 142 SH3.  
 FT DOMAIN 148 245 SH2.  
 FT DOMAIN 267 520 SH3.  
 FT DOMAIN 270 523 PROTEIN KINASE.  
 FT NP\_BIND 270 523 ATP (BY PKC) (BY SIMILARITY).  
 FT BINDING 276 284 ATP (BY SIMILARITY).  
 FT ACT\_SITE 298 298 ATP (BY SIMILARITY).  
 FT MOD\_RES 389 389 BY SIMILARITY.  
 FT MOD\_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 SQ SEQUENCE 536 AA; 60715 MW; 87236822 CRC32;

Query Match 50.0%; Score 40; DB 1; Length 536;  
 Best Local Similarity 61.5%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SCGTRHSTGCT 14  
 :||:|||||  
 Db 72 TGLTRRGCTGVT 84

RESULT 12  
 FYN\_XIPHE STANDARD: PRT: 536 AA.  
 AC P27446:  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE PROTO-ONCOGENE TYROSINE-PROTEIN KINASE FYN (EC 2.7.1.112) (P59-FYN).  
 GN FYN.  
 OS Xiphophorus helleri.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Atherinomorpha;  
 OC Cyprinodontiformes; Cyprinodontoidae; Poeciliidae; Xiphophorus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-RIO LANCETILLA;  
 RX MEDLINE: 91187435.  
 RA HANNIG G., OTTILE S., SCHARL M.;  
 RT "Conservation of structure and expression of the c-yes and fyn genes  
 in lower vertebrates.";  
 RL Oncogene 6:361-369(1991).  
 CC -1- FUNCTION: IMPLICATED IN THE CONTROL OF CELL GROWTH.  
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +  
 CC PROTEIN TYROSINE PHOSPHATE.  
 CC -1- SUBUNIT: ASSOCIATES THROUGH ITS SH3 DOMAIN, TO THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: TO OTHER PROTEIN-TYROSINE KINASES IN THE CATALYTIC  
 CC DOMAIN, BELONGS TO THE SRC SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X54971; CA38715.1;  
 DR HSSP: P06241; 1AOT.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PFAM: PF00069; PKinase; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS50001; SH2; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 DR Proto-oncogene; Tyrosine-protein kinase; Phosphorylation;  
 KW ATP-binding; Myristylation; SH3 domain; SH2 domain; Palmitate;  
 KW Lipoprotein.  
 KW INIT\_MET 0  
 FT LIPID 1 0  
 FT LIPID 2 1 MYRISTATE (BY SIMILARITY).  
 FT LIPID 5 5 PALMITATE (BY SIMILARITY).  
 FT LIPID 5 5 PALMITATE (BY SIMILARITY).  
 FT DOMAIN 81 142 SH3.  
 FT DOMAIN 148 245 PROTEIN KINASE.  
 FT DOMAIN 270 523 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 FT MOD\_RES 11 11 ATP (BY SIMILARITY).  
 FT NP\_BIND 276 284 ATP (BY SIMILARITY).  
 FT BINDING 298 389 ATP (BY SIMILARITY).  
 FT ACT\_SITE 389 389 BY SIMILARITY.  
 FT MOD\_RES 419 419 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 530 530 PHOSPHORYLATION (BY SIMILARITY).  
 SO SEQUENCE 536 AA; 60316 MW; BA03DB12 CRC32;

Query Match 50.0%; Score 40; DB 1; Length 536;  
 Best Local Similarity 51.5%; Pred. No. 10;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 SCGTRHSTGCT 14  
 :||:|||||  
 Db 72 TGLTRRGCTGVT 84

RESULT 13  
 MASD\_SCHUA STANDARD: PRT: 120 AA.  
 ID MASD\_SCHUA  
 AC P13411:  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE IMMUNOGENIC MIRACIDIAL ANTIGEN 5D (FRAGMENT).  
 GN 5D.  
 OS Schistosoma japonicum (Blood fluke).  
 OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeididae;  
 OC Schistosomatidae; Schistosomatidae; Schistosoma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-PHILIPPINE;  
 RX MEDLINE: 89261924.  
 RA SCALDON B.J., BOGITSH B.J., CARTER C.E.;  
 RT "Characterization of a large gene family in Schistosoma japonicum  
 that encodes an immunogenic miracidial antigen.";  
 RL Mol. Biochem. Parasitol. 33:105-112(1989).  
 CC -1- DEVELOPMENTAL STAGE: MIRACIDIA.  
 CC -1- SIMILARITY: NEARLY IDENTICAL TO IMMUNOGENIC MIRACIDIAL  
 CC ANTIGENS 8I, 8I AND 8C.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M25214; AAA29851.1;  
 DR Antigen; Multigene family.  
 KW Antigen; Multigene family.  
 FT NON\_TER 1  
 SO SEQUENCE 120 AA; 13541 MW; C5BE9039 CRC32;

Query Match 50.0%; Score 40; DB 1; Length 120;  
 Best Local Similarity 57.1%; Pred. No. 2.1;  
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 QSGTRHSTGCT 14  
 :||:|||||  
 Db 77 QGGRNRHGGCT 90

RESULT 14  
 RPBL\_EUPOC STANDARD: PRT: 478 AA.  
 ID RPBL\_EUPOC  
 AC P28364:  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (EC 2.7.7.6)  
 GN (FRAGMENT).  
 OS Euplotes octocarinatus.  
 OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Euplotidae; Euplotes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-(68)-VIII;  
 RX MEDLINE: 93027138.





Sat Jan 15 11:45:08 2000

us-08-991-628-7.rsp

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:24 ; Search time 59.21 Seconds  
(without alignments)  
15.518 Million cell updates/sec

Title: US-08-991-628-7  
Perfect score: 80  
Sequence: 1 QSGTMRHSTGCTN 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL\_11.\*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	69	86.2	993 11 Q35902	Q35902 mus musculus
2	44	55.0	421 0	Q86590 streptomyces
3	43	53.8	1597 5 Q61346	Q61346 drosophila
4	41	51.2	572 12 Q65689	Q65689 bovine para
5	40	50.0	534 4 Q16248	Q16248 homo sapien
6	40	50.0	584 5 Q25289	Q25289 leishmania
7	40	50.0	1921 5 Q01349	Q01349 drosophila
8	40	50.0	537 11 Q62844	Q62844 tritus norv
9	40	50.0	69 12 P90363	P90363 human cytom
10	40	50.0	69 12 P90292	P90292 human cytom
11	40	50.0	83 12 Q68570	Q68570 human cytom
12	40	50.0	171 12 P89538	P89538 human cytom
13	40	50.0	170 12 P90392	P90392 human cytom
14	40	50.0	169 12 P90393	P90393 human cytom
15	40	50.0	172 12 P90394	P90394 human cytom
16	40	50.0	718 12 Q69174	Q69174 human cytom
17	40	50.0	905 12 Q69173	Q69173 human cytom
18	39	48.8	275 2 Q44378	Q44378 agrobacteri
19	39	48.8	1879 2 P72938	P72938 synechocyst
20	39	48.8	797 5 Q44536	Q44536 caenorhabdi
21	38	47.5	1089 2 Q07800	Q07800 mycobacteri
22	38	47.5	137 2 P74924	P74924 thermotoga
23	38	47.5	425 5 Q44433	Q44433 drosophila
24	38	47.5	1060 5 Q16926	Q16926 aedes aegypt
25	38	47.5	396 5 Q22250	Q22250 caenorhabdi

## ALIGNMENTS

RESULT 1  
ID Q35902 PRELIMINARY; PRT; 993 AA.  
AC Q35902;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE DESMOGLEIN 3 (FRAGMENT).  
GN DSG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C.  
RA ISHIKAWA H., Li K., UETTO J.;  
Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; U86016; AAB65091.1; -  
DR PIRAF; P00028; cadherin.4;  
DR PROSITE; PS00232; CADHERIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON\_TER 993  
SQ SEQUENCE 993 AA; 10788 MW; 881794BD CRC32;

Query Match 86.2%; Score 69; DB 11; Length 993;  
Best local similarity 100.0%; Pred. No. 0.00051;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SGTMRHSTGCT 14  
DB 743 SGTMRHSTGCT 755  
RESULT 2  
ID Q86590 PRELIMINARY; PRT; 421 AA.  
AC Q86590;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE PUTATIVE SECRETED PROTEIN.  
GN SC2H4.07C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA OLIVER K., HARRIS D.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RA PARKILL J., BARRELL B.G., RAJANDREAM M.A.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-A3(2);  
 RX MEDLINE: 97000351.  
 RA RDEMBACH M., KIESER H.M., DENAPATTE D., EICHNER A., CULLUM J.,  
 RA KINASHI H., HOPWOOD D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 the 8 Mb streptomyces coelicolor A3(2) chromosome.";  
 RL Mol. Microbiol. 21:77-96(1996).  
 DR EMBL: AL031514; CAA20601.1;  
 SQ SEQUENCE 421 AA; 44174 MW; 7B444909 CRC32;

Query Match 55.0%; Score 44; DB 2; Length 421;  
 Best Local Similarity 69.2%; Pred. No. 4.4;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 SGTMRHSTGCT 14  
 ||:|||||  
 DB 48 GSGMRTRDIDGCT 60

RESULT 3  
 ID 061346 PRELIMINARY; PRT; 1597 AA.  
 AC 061346;  
 DT 01-AUG-1998 (TRENBLREL. 07, Created)  
 DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)  
 DT 01-MAY-1999 (TRENBLREL. 10, Last annotation update)  
 DE SMALL OPTIC LOBES.  
 GN SOL.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTONS;  
 RX MEDLINE: 96129280.  
 RA DE COUET H.G., FONG K.S., WEEDS A.G., MCLAUGHLIN P.J., MIKLOS G.L.;  
 RT "Molecular and mutational analysis of a gelsolin-family member  
 encoded by the flightless I gene of Drosophila melanogaster.";  
 RL Genetics 141:1049-1059(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTONS;  
 RX MEDLINE: 97289742.  
 RA MIKLOS G.L., YAMAMOTO M., BURNS R.G., MALESZKA R.;  
 RT "An essential cell division gene of Drosophila, absent from  
 Saccharomyces, encodes an unusual protein with tubulin-like and  
 myosin-like peptide motifs.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:5189-5194(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CANTONS;  
 RA MALESZKA R., DE COUET H.G., MIKLOS G.L.G.;  
 RT Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1998).  
 DR EMBL: AF017777; AAC28409.1;  
 DR FLYBASE: FBgn0003464; sol.  
 DR PFAM: PF00648; Peptidase\_C2; 1.  
 DR PFAM: PF00641; zf-RanBP; 6.  
 SQ SEQUENCE 1597 AA; 174696 MW; C6C4952E CRC32;

Query Match 53.8%; Score 43; DB 5; Length 1597;  
 Best Local Similarity 61.5%; Pred. No. 25;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 SGTMRHSTGCT 14  
 ||:|||||  
 DB 873 SGAIPIRHSTGGG 885

RESULT 4  
 ID 065689 PRELIMINARY; PRT; 572 AA.  
 AC 065689;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)  
 DE HEMAGGLUTININ-NEURAMINIDASE PROTEIN.  
 GN HN.  
 OS Bovine parainfluenza 3 virus.  
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;  
 OC Paramyxoviridae; Paramyxovirinae; Paramyxovirus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95287485.  
 RA BREKER-KLASSEN M.M., YOO D., MITTAL S.K., SORDEN S.D., HAINES D.M.,  
 RA BABIUK L.A.;  
 RT "Recombinant type 5 adenoviruses expressing bovine parainfluenza  
 RT virus type 3 glycoproteins protect Sigmund hispidus cotton rats from  
 RT bovine parainfluenza virus type 3 infection.";  
 RL J. Virol. 69:4308-4315(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA BREKER-KLASSEN M.M.;  
 RL Thesis (1995), Veterinary Infectious Disease Organization,  
 DR EMBL: U31671; AAB03692.1;  
 DR PFAM: PF00423; HN; 1.  
 SQ SEQUENCE 572 AA; 64668 MW; A16219EB CRC32;

Query Match 51.2%; Score 41; DB 12; Length 572;  
 Best Local Similarity 61.5%; Pred. No. 20;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 GTMRHSTGCTN 15  
 ||:|||||  
 DB 18 GTTRDRHSSKATN 30

RESULT 5  
 ID 016248 PRELIMINARY; PRT; 534 AA.  
 AC 016248;  
 DT 01-NOV-1996 (TRENBLREL. 01, Created)  
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)  
 DT 01-MAY-1999 (TRENBLREL. 10, Last annotation update)  
 DE PS9EYRN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95123064.  
 RA RIGLEY K., SLOCUMBE P., PROUDFOOT K., WAHID S., MANDAIR K.,  
 RA BEBBINGTON C.;  
 RT "Human p59fyn(T) regulates OKR3-induced calcium influx by a mechanism  
 RT distinct from p12 hydrolysis in Jurkat T cells.";  
 RL J. Immunol. 154:1136-1145(1995).  
 DR EMBL: S74774; AAB33113.1;  
 DR PFAM: PF00069; Pkinase; 1.  
 DR PFAM: PF00017; SH2; 1.  
 DR PFAM: PF00018; SH3; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
SQ SEQUENCE 534 AA; 60226 MW; D3941652 CRC32;

Query Match  
Best Local Similarity 50.0%; Score 40; DB 4; Length 534;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
OY 2 SGTMRHSTG 14  
DB 73 TGLTRRGTVT 85

RESULT 6  
025289 PRELIMINARY; PRT; 584 AA.  
AC 025289;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE MAJOR SURFACE GLYCOPROTEIN PRECURSOR.  
GN GP63.  
OS Leishmania guyanensis.  
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-MHOM/BR/75/M4147A;  
RX MEDLINE: 94187792.  
RA STEINKRAUS H.B., GREER J.M., STEPHENSON D.C., LANGER P.J.;  
RT "Sequence heterogeneity and polymorphic gene arrangements of the"  
RL Leishmania guyanensis gp63 gene."  
RL Mol. Biochem. Parasitol. 62:173-185(1993).  
DR EMBL: I16777; AAA29239.1.  
DR PFAM: PF01457; Peptidase\_M8; 1.  
KW Signal.  
FT CHAIN 1 37 POTENTIAL.  
FT SIGNAL 99 577 MAJOR SURFACE GLYCOPROTEIN.  
SQ SEQUENCE 584 AA; 63896 MW; 6AA97A3E CRC32;

Query Match  
Best Local Similarity 50.0%; Score 40; DB 5; Length 584;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 3 GTMRHSTG 12  
DB 491 GTRPRHSTG 500

RESULT 7  
001349 PRELIMINARY; PRT; 1921 AA.  
AC 001349;  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE KINESIN-73.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE: 97188425.  
RX LI H.P., LIU Z.M., NIRENBERG M.;  
RT "Kinesin-73 in the nervous system of Drosophila embryos."  
RL Proc. Natl. Acad. Sci. U.S.A. 94:1086-1091(1997).  
DR EMBL: U81788; AAB50404.1;  
DR FLYBASE: FBgn0019968; Khc-73.  
DR PFAM: PF01302; CAP\_GLY; 1.  
DR PFAM: PF00225; Kinesin; 1.  
DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
DR PROSITE; PS00845; CAP\_GLY; 1.

KW Motor protein; Microtubules; ATP-binding; Coiled coil.  
SQ SEQUENCE 1921 AA; 215047 MW; BF55A112 CRC32;

Query Match  
Best Local Similarity 50.0%; Score 40; DB 5; Length 1921;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 3 GTMRHSTG 13  
DB 1432 GVIRSHSTG 1442

RESULT 8  
062844 PRELIMINARY; PRT; 537 AA.  
AC 062844;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE PROTO-ONCOGENE FYN.  
GN P59FYN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SHRP STROKE-PRONE SPONTANEOUSLY HYPERTENSIVE;  
RC TISSUE-WHOLE BRAIN;  
RA NEMOTO K., SEKIMOTO M., KAGEYAMA H., FUKAMACHI K., NEMOTO F.,  
RA UENAGA T., SENNA E., TOMITA I.;  
RL Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.  
DR EMBL: U3365; AAA82942.1;  
DR PFAM: PF00069; phosphatase; 1.  
DR PFAM: PF00017; SH2; 1.  
DR PFAM: PF00018; SH3; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
SQ SEQUENCE 537 AA; 60701 MW; D0A1DD46 CRC32;

Query Match  
Best Local Similarity 50.0%; Score 40; DB 11; Length 537;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
OY 2 SGTMRHSTG 14  
DB 73 TGLTRRGTVT 85

RESULT 9  
P90363 PRELIMINARY; PRT; 69 AA.  
AC P90363;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
DE GLYCOPROTEIN B (FRAGMENT).  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERLIN H;  
RA BINDER T., STEGERT W., SCHMIDT C.A.;  
RL Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.  
DR EMBL: X99848; CAA6158.1;  
DR NON\_TER 1 1  
FT NON\_TER 69 69  
SQ SEQUENCE 69 AA; 7905 MW; F68398B7 CRC32;

Query Match  
50.0%; Score 40; DB 12; Length 69;

Best Local Similarity 80.0%; Pred. No. 3.5;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRSTGCTN 15  
||| ||| |||  
Db 25 RTRSTGCTN 34

## RESULT 10

ID P90292 PRELIMINARY; PRT; 69 AA.  
AC P90292;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
DE GLYCOPROTEIN B (FRAGMENT).  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERLIN E;  
RA BINDER T., SIEBERT W., SCHMIDT C.A.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL, X99846; CAA68156.1; -.  
FT NON\_TER 1  
FT SEQUENCE 69 AA; 7905 MW; F68398B7 CRC32;

Query Match 50.0%; Score 40; DB 12; Length 69;  
Best Local Similarity 80.0%; Pred. No. 3.5;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRSTGCTN 15  
||| ||| |||  
Db 25 RTRSTGCTN 34

## RESULT 11

ID 068570 PRELIMINARY; PRT; 83 AA.  
AC 068570;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE GLYCOPROTEIN B (FRAGMENT).  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-IV.1;  
RX MEDLINE; 96365443.  
RA LASRY S., DENV P., ASSILOT C., RAUZY M., BOUCHER J., GUYOT C.,  
RA LEROUX M.C., LIAKROMSKI A., REINERT P., NICOLAS J.C.,  
RT "Interstrain variations in the cytomegalovirus (CMV) glycoprotein B  
RT gene sequence among CMV-infected children attending six day care  
RT centers.";  
RL J. Infect. Dis. 174:606-609 (1996).  
DR EMBL; U52134; AAB53249.1; -.  
FT NON\_TER 1  
FT SEQUENCE 83 AA; 9638 MW; 72E2051D CRC32;

Query Match 50.0%; Score 40; DB 12; Length 83;  
Best Local Similarity 80.0%; Pred. No. 4.3;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRSTGCTN 15  
||| ||| |||  
Db 29 RTRSTGCTN 38

## RESULT 12

ID P89538 PRELIMINARY; PRT; 171 AA.  
AC P89538;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).  
GN GP55.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1553-91;  
RX MEDLINE; 96262011.  
RA SHEP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,  
RA PERGOLIZZI R.;  
RT "Cytomegalovirus glycoprotein B groups associated with retinitis in  
RT AIDS.";  
RL J. Infect. Dis. 174:184-187 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1553-91;  
RA SHEP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88701; AAB48525.1; -.  
FT NON\_TER 1  
FT NON\_TER 171  
FT SEQUENCE 171 AA; 19461 MW; CB157571 CRC32;

Query Match 50.0%; Score 40; DB 12; Length 171;  
Best Local Similarity 80.0%; Pred. No. 8.8;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRSTGCTN 15  
||| ||| |||  
Db 97 RTRSTGCTN 106

## RESULT 13

ID P90392 PRELIMINARY; PRT; 170 AA.  
AC P90392;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).  
GN GP55.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-419-91;  
RX MEDLINE; 96262011.  
RA SHEP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,  
RA PERGOLIZZI R.;  
RT "Cytomegalovirus glycoprotein B groups associated with retinitis in  
RT AIDS.";  
RL J. Infect. Dis. 174:184-187 (1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-419-91;  
RA SHEP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88700; AAB48524.1; -.  
FT NON\_TER 1  
FT NON\_TER 170  
FT SEQUENCE 170 AA; 19413 MW; FEOBE0AD CRC32;

## Query Match

Best Local Similarity 50.0%; Score 40; DB 12; Length 170;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRHSTGCTN 15  
DB 97 RTRRSTGCTN 106

## RESULT 14

P90393 PRELIMINARY; PRT; 169 AA.

AC P90393;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
DE GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).  
GN GP55.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1712-91;  
RX MEDLINE; 96262011.  
RA SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,  
RA PERGOLIZZI R.;  
RT "Cytomegalovirus glycoprotein B groups associated with retinitis in  
RT AIDS.";  
RL J. Infect. Dis. 174:184-187(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-1712-91.  
RA SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88703; AAB48527.1; -.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 169 AA; 19284 MW; C8D31BF5 CRC32;

## Query Match

Best Local Similarity 50.0%; Score 40; DB 12; Length 169;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRHSTGCTN 15  
DB 96 RTRRSTGCTN 105

## RESULT 15

P90394 PRELIMINARY; PRT; 172 AA.

AC P90394;  
DT 01-MAY-1997 (TREMblrel. 03, Created)  
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)  
DE GLYCOPROTEIN B VARIABLE REGION (FRAGMENT).  
GN GP55.  
OS Human cytomegalovirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae; Cytomegalovirus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-2515-92;  
RX MEDLINE; 96262011.  
RA SHEPP D.H., MATCH M.E., ASHRAF A.B., LIPSON S.M., MILLAN C.,  
RA PERGOLIZZI R.;  
RT "Cytomegalovirus glycoprotein B groups associated with retinitis in  
RT AIDS.";  
RL J. Infect. Dis. 174:184-187(1996).

## [2]

RP SEQUENCE FROM N.A.  
RC STRAIN-2515-92;  
RA SHEPP D.H., MATCH M.E., LIPSON S.M., MILLAN C., PERGOLIZZI R.G.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U88704; AAB48528.1; -.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 172 AA; 19597 MW; 5DC159A7 CRC32;

## Query Match

Best Local Similarity 50.0%; Score 40; DB 12; Length 172;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 6 RTRHSTGCTN 15  
DB 97 RTRRSTGCTN 106

Search completed: January 12, 2000, 23:15:25  
Job time: 188 sec

Sat Jan 15 11:45:08 2000

us-08-991-628-7.rspt

Page 6



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:38 : Search time 28.55 seconds

(without alignments)  
15.282 Million cell updates/sec

Title: US-08-991-628-4

Sequence: 1 TPMFLSRNNGEVRT 15

Scoring table: BIOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38:\*

Word size: 0

Number of hits that pass the threshold: 80000

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	76	100.0	999	1 DSG3_HUMAN	P39296 homo sapien
2	64	84.2	1049	2 DSG1_HUMAN	Q02413 homo sapien
3	56	73.7	1043	1 DSG1_BOVIN	Q03763 bos taurus
4	43	56.6	1117	1 DSG2_HUMAN	Q14126 homo sapien
5	40	52.6	571	1 ASN1_YEAST	P4089 saccharomyc
6	40	52.6	571	1 ASN2_YEAST	P43090 saccharomyc
7	40	52.6	375	1 C15Y_MYCSE	P26491 mycobacteri
8	38	50.0	585	1 ASN1_LOTJA	P49092 lotus japon
9	38	50.0	585	1 ASN1_PEA	P19251 lotus japon
10	38	50.0	585	1 ASN2_LOTJA	P49093 lotus japon
11	38	50.0	582	1 ASN2_PEA	P19252 lotus japon
12	38	50.0	583	1 ASN3_ARATH	P43078 arabidopsis
13	38	50.0	583	1 ASN3_BRAOL	P43091 brassica ol
14	38	50.0	585	1 ASN3_MAIZE	P43094 zea mays (m
15	38	50.0	590	1 ASN3_ORYSA	Q43011 oryza sativ
16	38	50.0	524	1 ASN3_SAMAU	Q24338 sandersonia
17	38	50.0	585	1 ASN3_RRITS	Q24661 triphysaria
18	37	48.7	186	1 ATPD_RHOB	P05437 rhodopsin
19	37	48.7	389	1 C15Z_ECOLI	P31660 escherichia
20	37	48.7	389	1 C15Z_SALTY	Q56063 escherichia
21	37	48.7	426	1 ENO_HELPY	P48285 helicobacte
22	37	48.7	576	1 RECN_BACSU	P17894 bacillus su
23	37	48.7	108	1 Y388_MYCN	P73209 mycoplasma
24	36	47.4	882	1 CAD1_HUMAN	P1830 homo sapien
25	36	47.4	790	1 CAD6_HUMAN	P55285 homo sapien
26	36	47.4	789	1 CAD6_RAT	P55280 rattus norv
27	36	47.4	730	1 CAD6_MOUSE	P33146 mus musculu
28	36	47.4	369	1 MURB_MYCTU	Q11148 mycobacteri
29	36	47.4	733	1 SUP_DROME	P25891 drosophila
30	36	47.4	152	1 V66_HPV03	P36799 human papil
31	35	46.1	1314	1 ADRE_YEAST	P09547 saccharomyc
32	35	46.1	1286	1 AIDA_ECOLI	Q03155 escherichia
33	35	46.1	554	1 ASNB_ECOLI	P22106 escherichia
34	35	46.1	589	1 ASNS_ASPOF	P31752 aspergillus o
35	35	46.1	379	1 C15Y_ABDG2	Q34002 antarctic b
36	35	46.1	396	1 CYB_PETNA	Q35534 petromyzon
37	35	46.1	488	1 DSBG_ECOLI	P36655 petromyzon
38	35	46.1	5147	1 FAT_DROME	P34450 drosophila
39	35	46.1	609	1 HAPT_VIBCH	P24153 vibrio chol

## ALIGNMENTS

RESULT	1	STANDARD	PRT	999 AA.
DSG3_HUMAN				
ID	DSG3_HUMAN			
AC	P39296			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MARGAI W., KLAUS-KOYTUN V., STANLEY J.R.;			
RA	"Autoantibodies against a novel epithelial cadherin in pemphigus			
RT	vulgaris, a disease of cell adhesion."			
RL	Cell 67:869-877(1991).			
CC	-1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.			
CC	INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE			
CC	FILAMENTS MEDIATING CELL-CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL, ESOPHAGUS AND			
CC	CARCINOMAS.			
CC	-1- DOMAIN: CALDUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS			
CC	(POTENTIAL).			
CC	-1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN			
CC	DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE			
CC	LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES			
CC	AGAINST DSG3.			
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; M76482; AAA60230.1; -			
DR	PIR; A41088; IJHUG3.			
DR	HSSP; P09803; IEDH.			
DR	MIM; 169615; -			
DR	PFAM; PF00028; cadherin.4.			
DR	PROSITE; PS00232; CADHERIN.3.			
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;			
KW	Calcium-binding; Repeat.			
FT	SIGNAL	1	23	POTENTIAL.
FT	PROPEP	24	49	POTENTIAL.
FT	CHAIN	50	99	DESMOGLEIN 3.
FT	DOMAIN	50	615	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	616	640	POTENTIAL.
FT	DOMAIN	641	999	CYTOSOLASMIC (POTENTIAL).
FT	REPEAT	50	158	CADHERIN 1.
FT	REPEAT	159	268	CADHERIN 2.
FT	REPEAT	269	383	CADHERIN 3.
FT	REPEAT	386	499	CADHERIN 4.
FT	REPEAT	910	935	DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966	DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110	110	POTENTIAL.

P30015 escherichia  
P04675 bradyrhizob  
P30760 mycobacteri  
P47766 mycobacteri  
Q26278 methanobact  
Q10327 schizosacch

Query Match  
Best Local Similarity 100.0%; Score 76; DB 1; Length 999;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFFLSRNTGEVRT 15  
DB 206 TPFFLSRNTGEVRT 220

RESULT 2  
DSGL\_HUMAN STANDARD; PRT: 1049 AA.

AC 002413;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
GN DSG1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KERATINOCYTES;  
RX MEDLINE: 91271279.  
RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALOTIS P., POYNTER D.,  
RA ARNEJANN J., RUTMAN A.J., PIDSELEY S.C., WATT F.M., REES D.A.,  
RA BUXTON R.S., MAGEE A.I.;  
RT "Desmosomal glycoprotein DGL, a component of intercellular desmosome  
junctions, is related to the cadherin family of cell adhesion  
molecules".  
RL Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).  
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.  
CC -1- DOMAIN: CALCIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
(POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X56654; CA39976.1; --  
CC DR PIR: S16906; ICHDGL.  
CC DR HSSP: P09803; LEDH.  
CC DR MIM: 125670; --  
CC PRAM: PF00028; cadherin. 4.  
CC KW PROSITE: PS00232; CADHERIN. 2.  
CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
CC Calcium-binding; Repeat.  
CC FT SIGNAL 1 23 POTENTIAL.  
CC FT PROPEP 24 49 POTENTIAL.  
CC FT CHAIN 50 1049 DESMOGLEIN 1.  
CC FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).  
CC FT TRANSLEM 546 570 POTENTIAL.  
CC FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).  
CC FT REPEAT 50 158 CADHERIN 1.  
CC FT REPEAT 159 270 CADHERIN 2.  
CC FT REPEAT 271 385 CADHERIN 3.  
CC FT REPEAT 386 497 CADHERIN 4.  
CC

Query Match  
Best Local Similarity 66.7%; Score 64; DB 1; Length 1049;  
Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFFLSRNTGEVRT 15  
DB 206 SPFFLNNTGEVRT 220

RESULT 3  
DSGL\_BOVIN STANDARD; PRT: 1043 AA.

AC 003763;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
GN DSG1.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
OC Bovinae; Bos.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-MUZZLE EPITHELIUM;  
RX KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;  
RX Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 44-1043 FROM N.A.  
RC TISSUE-MUZZLE EPITHELIUM;  
RX MEDLINE: 9116865.  
RA KOCH P.J., WALSH M.J., SCHWELZ M., GOLDSCHMIDT M.D.,  
RA ZIMBELMANN R., FRANK W.W.;  
RT "Identification of desmoglein, a constitutive desmosomal  
glycoprotein, as a member of the cadherin family of cell adhesion  
molecules".  
RL Eur. J. Cell Biol. 53:1-12(1990).  
RN [3]  
RP REVISIONS, AND SEQUENCE OF 101-123.  
RX MEDLINE: 92037656.  
RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
RA FRANK W.W.;  
RT "Complete amino acid sequence of the epidermal desmoglein precursor  
polypeptide and identification of a second type of desmoglein gene".  
RT Eur. J. Cell Biol. 55:200-208(1991).  
RN [4]  
RP SEQUENCE OF 44-493 FROM N.A.  
RX MEDLINE: 91097553.  
RA GOODMAN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COMIN P.;  
RT "Desmoglein shows extensive homology to the cadherin family of cell  
adhesion molecules".  
RL Biochem. Res. Commun. 173:1224-1230(1990).  
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.  
CC -1- DOMAIN: CALCIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
(POTENTIAL).  
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: X58466; CAA1380.1; -  
 DR EMBL: X57784; CAA40930.1; -  
 DR EMBL: M58165; AAA62709.1; -  
 DR PIR: S14603; IJBOG1.  
 DR HSSP: P09803; 1EDH.  
 DR PFAM: PF00028; cadherin; 3.  
 DR PROSITE: PS00232; CADHERIN; 2.  
 KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 KM Calcium-binding; Repeat.  
 FT SIGNAL 1 23  
 FT PROPEP 24 49  
 FT CHAIN 50 1043  
 FT DOMAIN 50 548  
 FT TRANSMEM 549 573  
 FT DOMAIN 574 1043  
 FT REPEAT 159 158  
 FT REPEAT 159 270  
 FT REPEAT 271 385  
 FT REPEAT 386 498  
 FT REPEAT 498 845  
 FT REPEAT 845 875  
 FT REPEAT 875 905  
 FT REPEAT 906 933  
 FT REPEAT 934 962  
 FT DOMAIN 963 1012  
 FT CARBOHYD 110 110  
 FT CARBOHYD 180 180  
 FT CARBOHYD 496 496  
 FT CARBOHYD 124 124  
 FT CONFLICT 124 124  
 FT SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;

Query Match 73.7%; Score 56; DB 1; Length 1043;  
 Best Local Similarity 60.0%; Pred. No. 0.018;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 TPFLLSRNTEVPT 15  
 DB 206 SPFFLLNRYTGEIRT 220

RESULT 4  
 DSG2\_HUMAN STANDARD; PRT; 1117 AA.  
 AC Q14126;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 2 PRECURSOR (HDC).  
 GN DSG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-COLON CARCINOMA;  
 RA MEDLINE; 94192736.  
 RA SCHAEFER S., KOCH P.J., FRANK W.W.;  
 RT "Identification of the ubiquitous human desmoglein, Dsg2, and the  
 RT expression catalogue of the desmoglein subfamily of desmosomal  
 RT cadherins."  
 RL Exp. Cell Res. 211:391-399(1994).  
 RN [2]  
 RP SEQUENCE OF 777-1117 FROM N.A.  
 RX MEDLINE; 92037656.

RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMMELMANN R., FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT peptide and identification of a second type of desmoglein gene";  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z26317; CAA81226.1; -  
 DR HSSP: P15116; INCI.  
 DR MIM: 125671; -  
 DR PFAM: PF00028; cadherin; 4.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;  
 KM Cytoskeleton; Calcium-binding.  
 FT SIGNAL 1 23  
 FT PROPEP 24 48  
 FT CHAIN 49 1117  
 FT DOMAIN 49 608  
 FT TRANSMEM 609 633  
 FT DOMAIN 634 1117  
 FT REPEAT 159 159  
 FT REPEAT 160 272  
 FT REPEAT 273 387  
 FT REPEAT 388 502  
 FT REPEAT 502 911  
 FT REPEAT 912 941  
 FT REPEAT 942 967  
 FT REPEAT 968 991  
 FT REPEAT 992 1020  
 FT REPEAT 1021 1050  
 FT CARBOHYD 111 111  
 FT CARBOHYD 181 181  
 FT CARBOHYD 308 308  
 FT CARBOHYD 461 461  
 FT CARBOHYD 513 513  
 FT SEQUENCE 1117 AA; 122385 MW; 8403898 CRC32;

Query Match 56.6%; Score 43; DB 1; Length 1117;  
 Best Local Similarity 50.0%; Pred. No. 4.8;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 2 PMFLLSRNTEVPT 15  
 DB 208 PVFFLLNRYTGEIRT 221

RESULT 5  
 ASNL\_YEAST STANDARD; PRT; 571 AA.  
 ID ASNL\_YEAST  
 AC P49089;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYSING] 1 (EC 6.3.5.4)  
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 1).  
 GN ASNI OR YPR145W OR P9659.3.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;

```

RT of ty and three tRNA genes." ;
RL Yeast 13:171-176(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA VAN DYCK L., SKALA J., DE MERGIOSSE P., PURNELLE B., TALLA E.,
RA NAMROCKI A., DEL BINO S., GOFFEAT A.,
RA SMURROTT (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE = AMP +
CC PYRPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES
CC -----
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC the European Bioinformatics Institute. There are no restrictions on its use
CC by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/).
CC send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83099; CAAS8159.1; -.
DR EMBL; Z72809; GA97135.1; -.
DR SGD; L0003156; ASN2
DR PRFAM; PF00310; GATase_2; 1.
DR PFAM; PF00733; Asn_synthase; 1.
DR PROSITE; PS00443; GATASE_TYPE_II; 1.
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KM Multigene family.
FT INIT MET 0 BY SIMILARITY.
FT ACCT SITE 1 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 571 AA; 64461 MW; 926C9376 CRC32;
OY 1 TPMFLSLR 8
Db 345 TPMFLSLR 352
Query Match 52.6%; Score 40; DB 1; Length 571;
Best Local Similarity 100.0%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
RESULT 7
CISY_MYCSM STANDARD; PRT; 375 AA.
AC P26491;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE CITRATE SYNTHASE (EC 4.1.3.7).
GN GLTA.
OS Mycobacterium smegmatis.
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales;
OC Actinomycetiales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RA KOHN J.C.;
RA DAVID M., LOBINSKY-MINK S., BENZI A., SUSSA M., UNITZUR S.,
RX STRAIN-ATCC 607;
RX MEDLINE; 91354207.
RT "Clonate synthase from Mycobacterium smegmatis. Cloning, sequence
RT determination and expression in Escherichia coli.";
RL Blochem. J. 278:225-234(1991).
CC -1- CATALYTIC ACTIVITY: CITRATE + COA = ACETYL-COA + H(2)O +
CC OXALACETATE.
CC -1- ENZYME REGULATION: ALLOSTERICALLY INHIBITED BY NADH.
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.
CC -1- SUBUNIT: HOMOHEXAMER.
CC -1- MISCELLANEOUS: CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS
CC CAPABLE OF OXIDATIVE METABOLISM.
CC -1- SIMILARITY: BELONGS TO THE CITRATE SYNTHASE FAMILY.

```

```

KW PROSITE; PS00443; GATASE_TYPE.II; 1.
DR Ligase; Asparagine biosynthesis; Glutamine amidotransferase;
KM Multigene family.
FT INIT_MET 0 BY SIMILARITY.
FT ACT_SITE 1 GATASE (BY SIMILARITY).
SQ SEQUENCE 585 AA; 66330 MW; 7FDC0436 CRC32;

Query Match 50.0%; Score 38; DB 1; Length 585;
Best Local Similarity 87.5%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFELSR 8
DB 321 TPFELMSR 328

RESULT 9
ASN1_PEA STANDARD; PRT; 585 AA.
ID ASN1_PEA
AC P19231; O49925;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1991 (Rel. 38, Last annotation update)
DE ASPARAGINE SYNTHETASE, NODULE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)
DE [GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE].
GN AS1.
OS Pisum sativum (Garden pea).
OC Euarvota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Pisum.
OC [1]
OC SEQUENCE FROM N.A.
RN RA STRAIN-CV, SPARKLE; TISSUE=ROOT NODULES;
RX MEDLINE; 90151604.
RA TSAI F.Y., CORUZZI G.M.;
RT "dark-induced and organ-specific expression of two asparagine
RT synthetase genes in Pisum sativum.";
RL EMBO J. 9:323-332(1990).
RN [2]
RP SEQUENCE OF 1-83 FROM N.A.
RC STRAIN-CV, FELTHAM FIRST;
RA NGAI N., TSAI F.Y., CORUZZI G.M.;
RT "light-induced transcriptional repression of the pea AS1 gene:
RT identification of cis-elements and transactors.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +
CC PHOSPHOSPHATE + L-ASPARGINE + L-GLUTAMATE.
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.
CC -1- TISSUE SPECIFICITY: NODULE.
CC -1- INDUCTION: DARK-INDUCED.
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL; X52179; CAA36429.1; -
DR EMBL; Y13321; CAA73762.1; -
DR PIR; S11444; AJPANI.
DR HSSP; P17169; IGMS.
DR PFAM; PF00310; GATase_2; 1.
DR PFAM; PF00733; asn_synthase; 1.
DR PROSITE; PS00443; GATASE_TYPE.II; 1.
KW Ligase; asparagine biosynthesis; glutamine amidotransferase;
KM Multigene family.

```

FT INT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 585 AA: 66222 MW; 9AB4FBC5 CRC32;

Query Match 50.0%; Score 38; DB 1; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFLLSR 8  
 |||||:  
 DB 322 TPFLLMSR 329

RESULT 10  
 ID ASN2\_LOTJA STANDARD; PRT; 585 AA.  
 AC P49093;  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] 2 (EC 6.3.5.4)  
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE 2).  
 GN AS2.  
 OS Lotus japonicus.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 OC Lotus.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-CV, FELTHAM FIRST;  
 RX MEDLINE: 96270368.  
 RA WATERHOUSE R.N., SMYTH A.J., MASSONEAU A., PROSSER I.M.,  
 RA CLARKSON D.T.;  
 RT "Molecular cloning and characterisation of asparagine synthetase from  
 RT Lotus japonicus: dynamics of asparagine synthesis in N-sufficient  
 RT conditions";  
 RL Plant Mol. Biol. 30:883-897(1996)  
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +  
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
 CC AMIDORANFERASES.  
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X89410; CA61590.1; -  
 DR HSRP; P17169; IGMS.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PROSITE: PS00443; GATASE, TYPE II; 1.  
 KW Ligase; Asparagine biosynthesis; glutamine amidotransferase;  
 KW Multigene family.  
 FT INT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 585 AA: 65838 MW; 2FE40574 CRC32;

Query Match 50.0%; Score 38; DB 1; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TPFLLSR 8  
 |||||:  
 DB 321 TPFLLMSR 328

RESULT 11  
 ID ASN2\_PEA STANDARD; PRT; 582 AA.  
 AC P19252; 049926;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE ASPARAGINE SYNTHETASE, ROOT [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4)  
 DE (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).  
 GN AS2.  
 OS Pisum sativum (Garden pea).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 OC Pisum.  
 [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-CV, SPARKLE; TISSUE=ROOT;  
 RX MEDLINE: 90151604.  
 RA TSAI F.Y., CORUZZI G.M.;  
 RA "Dark-induced and organ-specific expression of two asparagine  
 RA synthetase genes in Pisum sativum";  
 RL EMBO J. 9:323-332(1990).  
 [2]  
 RN SEQUENCE OF 1-83 FROM N.A.  
 RC STRAIN-CV, FELTHAM FIRST;  
 RX NGAI N., TSAI F.Y., CORUZZI G.M.;  
 RT "Light-induced transcriptional repression of the pea As1 gene:  
 RT Identification of cis-elements and transactors";  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +  
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
 CC -1- TISSUE SPECIFICITY: ROOTS.  
 CC -1- INDUCTION: DARK-INDUCED.  
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
 CC AMIDORANFERASES.  
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X52180; CA36430.1; -  
 DR EMBL: Y13322; CA47363.1; -  
 DR PIR: S11443; AJPWN2.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PROSITE: PS00443; GATASE, TYPE II; 1.  
 KW Ligase; Asparagine biosynthesis; glutamine amidotransferase;  
 KW Multigene family.  
 FT INT\_MET 0 0 BY SIMILARITY.  
 FT ACT\_SITE 1 1 GATASE (BY SIMILARITY).  
 FT CONFLICT 33 33 E -> D (IN REF. 2).  
 FT CONFLICT 44 44 Y -> F (IN REF. 2).  
 FT CONFLICT 47 47 Q -> H (IN REF. 2).  
 FT CONFLICT 77 77 I -> L (IN REF. 2).  
 SQ SEQUENCE 582 AA: 65518 MW; 793421FA CRC32;

Query Match 50.0%; Score 38; DB 1; Length 582;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TPFLLSR 8  
 |||||:  
 DB 321 TPFLLMSR 328

RESULT 12  
ASNS\_ARATH STANDARD; PRT; 583 AA.  
AC ASNS\_ARATH  
AC P49078;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
RT 01-FEB-1996 (Rel. 33, Last annotation update)  
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).  
GN ASN1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Arabidopsis.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. COLUMBIA;  
RX MEDLINE: 95148732.  
RA LAM H.M., PENG S.S., CORUZZI G.M.;  
RT "Metabolic regulation of the gene encoding glutamine-dependent  
asparagine synthetase in Arabidopsis thaliana."  
PL Plant Physiol. 106:1347-1357(1994).  
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE - AMP +  
PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
AMIDOTRANSFERASES.  
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: L29083; AAA74359.1; -  
DR HSSP: P17169; IGMS.  
DR PFAM: PF00310; GATase-2; 1.  
DR PFAM: PF00733; Asn\_synthase; 1.  
DR PROSITE: PS00443; GATASE\_TYPE\_1; 1.  
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase;  
KM Multigene family.  
FT INIT MET 0 BY SIMILARITY.  
FT ACT SITE 1 GATASE (BY SIMILARITY).  
FT SEQUENCE 583 AA; 65489 MW; F8FC9672 CRC32;  
SQ

Query Match 50.0%; Score 38; DB 1; Length 583;  
Best Local Similarity 87.5%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLISR 8  
Db 321 TPEFLISR 328

RESULT 13  
ASNS\_BRAOL STANDARD; PRT; 585 AA.  
AC ASNS\_BRAOL  
AC P49091;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).  
OS Brassica oleracea (Caulliflower).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
OC

OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;  
OC Brassica.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. SHOGUN;  
RA DOWNS C.G., POGSON B.J., DAVIES K.M., ALMIRA E.C.;  
RT "An asparagine synthetase cDNA clone from Broccoli (Brassica oleracea  
L.)".  
PL Plant Gene Register PGR95-016.  
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE - AMP +  
PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
AMIDOTRANSFERASES.  
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X84448; CAA59138.1; -  
DR PFAM: PF00310; GATase-2; 1.  
DR PFAM: PF00733; Asn\_synthase; 1.  
DR PROSITE: PS00443; GATASE\_TYPE\_1; 1.  
KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase.  
KM Multigene family.  
FT INIT MET 0 BY SIMILARITY.  
FT ACT SITE 1 GATASE (BY SIMILARITY).  
FT SEQUENCE 585 AA; 65541 MW; B6DCFB50 CRC32;  
SQ

Query Match 50.0%; Score 38; DB 1; Length 585;  
Best Local Similarity 87.5%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLISR 8  
Db 322 TPEFLISR 329

RESULT 14  
ASNS\_MAIZE STANDARD; PRT; 585 AA.  
AC ASNS\_MAIZE  
AC P49094;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-DEPENDENT ASPARAGINE SYNTHETASE).  
GN ASN1 OR AS.  
OS Zea mays (maize).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
OC Poaceae; Zea.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CV. DEA; TISSUE-ROOT MERISTEM;  
RX MEDLINE: 96158342.  
RA CHEVALIER C., BOURGEOIS E., JUST D., RAYMOND P.;  
RT "Metabolic regulation of asparagine synthetase gene expression in  
maize (Zea mays L.) root tips".  
PL Plant J. 9:1-11(1996).  
CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARATE + L-GLUTAMINE - AMP +  
PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
AMIDOTRANSFERASES.  
CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL: X82849; CAA58052.1; -  
 DR MATZEDB: 79071; -  
 DR PFAM: PF00310; GATase\_2; 1.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PROSITE: PS00443; GATASE\_TYPE\_II; 1.  
 KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase.  
 FT INIT MET 0 BY SIMILARITY.  
 FT ACT SITE 1 1 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 585 AA: 66446 MW: 82480478 CRC32;

Query Match 50.0%; Score 38; DB 1; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 DB 321 TPMFLMSR 328

RESULT 15  
 ASNS\_ORYSA STANDARD; PRT; 590 AA.  
 ID ASNS\_ORYSA STANDARD; PRT; 590 AA.  
 AC 043011;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE ASPARAGINE SYNTHETASE [GLUTAMINE-HYDROLYZING] (EC 6.3.5.4) (GLUTAMINE-  
 DE DEPENDENT ASPARAGINE SYNTHETASE).  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;  
 OC Poaceae; Oryza.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, NIPPONBARE; TISSUE=CALLUS;  
 RA SUEYOSHI K., KAWACHI T., NAKAJIMA A., YAMAGATA H., SUGIMOTO T.,  
 RA IWASAKI T., OJI Y.;  
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +  
 CC PYROPHOSPHATE + L-ASPARAGINE + L-GLUTAMATE.  
 CC -1- PATHWAY: ASPARAGINE BIOSYNTHESIS.  
 CC -1- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-2 GLUTAMINE  
 CC AMIDOTRANSFERASES.  
 CC -1- SIMILARITY: TO OTHER ASN SYNTHETASES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U55873; AAB03991.1; -  
 DR PFAM: PF00310; GATase\_2; 1.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PROSITE: PS00443; GATASE\_TYPE\_II; 1.  
 KW Ligase; Asparagine biosynthesis; Glutamine amidotransferase.  
 FT INIT MET 0 BY SIMILARITY.  
 FT ACT SITE 1 1 GATASE (BY SIMILARITY).  
 SQ SEQUENCE 590 AA: 66096 MW: 3303930 CRC32;

Query Match 50.0%; Score 38; DB 1; Length 590;  
 Best Local Similarity 87.5%; Pred. No. 21;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TPMFLSR 8  
 DB 321 TPMFLMSR 328

Search completed: January 7, 2000, 13:25:39  
 Job time: 391 sec



Sat Jan 15 11:45:02 2000

us-08-991-628-4.rsp

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:41 ; Search time 65.16 Seconds  
(without alignments)  
10.858 Million cell updates/sec

Title: US-08-991-628-4  
Perfect score: 76  
Sequence: 1 TPFLLSNTGEVRT 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR\_62:\*

Word size: 0

Number of hits that pass the threshold: 142080

1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	76	100.0	999	1 IJHUG3	desmoglein 3 precu
2	64	84.2	1049	1 IJHUG1	desmoglein 1 precu
3	56	73.7	1043	1 IJBOG1	desmoglein 1 precu
4	43.5	57.2	295	2 S32892	hypothetical prote
5	43	56.6	1117	2 S38673	desmoglein 2 - hum
6	40	52.6	375	1 YKMY	citrate (sl)-synth
7	40	52.6	572	2 S52694	asparagine synthas
8	40	52.6	572	2 S55982	asparagine synthas
9	40	52.6	452	2 F69294	DNA repair protein
10	39	51.3	393	2 D70539	probable citrate s
11	39	51.3	163	2 A71536	hypothetical prote
12	39	51.3	302	2 C72756	probable multiple
13	38	50.0	583	1 AUPM2	asparagine synthas
14	38	50.0	586	1 AUPM1	asparagine synthas
15	38	50.0	586	2 S69182	asparagine synthas
16	38	50.0	586	2 S69183	asparagine synthas
17	38	50.0	586	2 S52387	asparagine synthas
18	38	50.0	591	2 T03602	probable asparagin
19	38	50.0	586	2 T02978	asparagine synthas
20	38	50.0	581	2 T08846	asparagine synthas
21	38	50.0	584	2 T12989	asparagine synthas
22	37	48.7	576	1 B35128	DNA repair and gen
23	37	48.7	186	2 S04671	H+-transporting AT
24	37	48.7	389	2 E64760	citrate (sl)-synth
25	37	48.7	426	2 S58684	phosphopyruvate hy
26	37	48.7	426	2 H71967	enolase - Helicoba
27	37	48.7	108	2 S73599	MG38 homolog D02-
28	37	48.7	367	2 T02000	hypothetical prote
29	37	48.7	852	2 B72685	hypothetical prote
30	37	48.7	365	2 E72475	hypothetical prote
31	37	48.7	211	2 E71334	hypothetical prote
32	37	48.7	593	2 T10301	ribonucleotide red
33	36	47.4	882	1 IJHUC1	cadherin 1 precurs
34	36	47.4	730	1 IJMSCM	M-cadherin - mouse
35	36	47.4	369	2 E70743	probable murb prot

36	47.4	790	2	I37016	cadherin-6 - human
37	47.4	789	2	I52701	K-cadherin - rat
38	47.4	790	2	I50178	cadherin-6B - Chic
39	47.4	483	2	T06459	62k sucrose-bindin
40	47.4	152	2	S36550	E6 protein - human
41	47.4	379	2	I49020	retinoid X recepto
42	47.4	484	2	I49018	retinoid X recepto
43	47.4	469	2	A56918	farneoid x-activa
44	47.4	466	2	H64904	hypothetical prote
45	47.4	1417	2	T00861	hypothetical prote

## ALIGNMENTS

```

RESULT 1
IJHUG3
desmoglein 3 precursor - human
N:Alternate names: pemphigus vulgaris antigen
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: A41088
R:Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.
Cell 67, 865-877, 1991
A:Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a
A:Reference number: A41088; MUID:92069753
A:Accession: A41088
A:Molecule type: mRNA
A:Residues: 1-999 <AMA>
A:Cross-references: GB:M76482; NID:9190751; PIDN:AAA60230.1; PID:9190752
C:Genetics:
A:Gene: GDB:DSG3
A:Cross-references: GDB:134030; OMIM:169615
A:Map position: 18q12.1-18q12.2
C:Superfamily: cadherin; cadherin repeat homology
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane
F:1-22/Domain: signal sequence #status predicted <SIG>
F:24-49/Domain: propeptide #status predicted <PRO>
F:50-999/Product: desmoglein homolog #status predicted <MAT>
F:50-615/Domain: extracellular #status predicted <EXT>
F:52-157/Domain: cadherin repeat homology <CR1>
F:160-267/Domain: cadherin repeat homology <CR2>
F:270-383/Domain: cadherin repeat homology <CR3>
F:390-495/Domain: cadherin repeat homology <CR4>
F:496-598/Domain: cadherin repeat homology <CR5>
F:616-635/Domain: transmembrane #status predicted <TM>
F:640-999/Domain: intracellular #status predicted <INT>
F:910-936/Domain: desmoglein repeat <DG1>
F:937-966/Domain: desmoglein repeat <DG2>
F:110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 76; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 7.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPFLLSNTGEVRT 15
DB 206 TPFLLSNTGEVRT 220

RESULT 2
IJHUG1
desmoglein 1 precursor - human
N:Alternate names: desmosomal glycoprotein I
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 22-Jun-1999
C:Accession: S16906; A39706; A61254; A61279; S16158
R:Buxton, R.S.
Submitted to the EMBL Data Library, November 1990
A:Reference number: S16906
A:Accession: S16906
A:Molecule type: mRNA

```

A:Residues: 1-1049 <BNX>  
 A:Cross-references: EMBL:X56654; NID:930505; PIDN:CAA9976.1; PID:930506  
 R:Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arnemann, J.; F  
 Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991  
 A:Title: Desmosomal glycoprotein Dg1, a component of intercellular desmosome junctions,  
 A:Reference number: A39706; MUID:912712729  
 A:Accession: A39706  
 A:Molecule type: mRNA  
 A:Residues: 24-1049 <WHE>  
 A:Cross-references: GB:X56654  
 R:Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.  
 J. Cell Sci. 99, 809-821, 1991  
 A:Title: Structural analysis and expression of human desmoglein: a cadherin-like compone  
 A:Reference number: A61254; MUID:92121251  
 A:Accession: A61254  
 A:Molecule type: mRNA  
 A:Residues: 26-1049 <NID>  
 R:Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Mage  
 Biochem. Soc. Trans. 19, 1060-1064, 1991  
 A:Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily  
 A:Reference number: A61279; MUID:92175187  
 A:Accession: A61279  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-55 <WH3>  
 C:Genetics:  
 A:Gene: GDB:DSGL  
 A:Cross-references: GDB:126563; OMIM:125670  
 A:Map position: 18q12.1-18q12.2  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-49/Domain: propeptide #status predicted <PRO>  
 F:50-149/Domain: extracellular #status predicted <EXT>  
 F:50-348/Domain: extracellular #status predicted <EXT>  
 F:160-269/Domain: extracellular #status predicted <EXT>  
 F:272-385/Domain: cadherin repeat homology <CR1>  
 F:392-493/Domain: cadherin repeat homology <CR2>  
 F:509-530/Domain: cadherin repeat homology <CR3>  
 F:549-569/Domain: transmembrane #status predicted <TM>  
 F:572-1049/Domain: intracellular #status predicted <INT>  
 F:840-869/Domain: desmoglein repeat <DG1>  
 F:870-899/Domain: desmoglein repeat <DG2>  
 F:900-927/Domain: desmoglein repeat <DG3>  
 F:928-956/Domain: desmoglein repeat <DG4>  
 F:969-1019/Domain: desmoglein repeat <DG5>  
 F:110/Binding site: glycyne/serine-rich  
 F:110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.2%; Score 64; DB 1; Length 1049;  
 Best Local Similarity 66.7%; Pred. No. 0.0013;  
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPEFLSRTGVEVT 15  
 Db 206 SPMTIRNTGEIRT 220  
 :||||:|||||

RESULT 3  
 IJBOG1  
 desmoglein 1 precursor - bovine  
 N:Alternate names: desmoglein BDGM  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
 C:Accession: S14603; A38872; A37785; S38721; A48173; S24412  
 R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.  
 Submitted to the EMBL Data Library, March 1991  
 A:Description: Complete sequence of the desmoglein precursor and evidence for the existe  
 A:Reference number: S14603  
 A:Accession: S14603  
 A:Molecule type: mRNA  
 A:Residues: 1-1043 <KOC>

A:Cross-references: EMBL:X58466; NID:9306; PIDN:CAA41380.1; PID:9307  
 R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.T.; Zimbelmann, R.; Franke, W.W.  
 Eur. J. Cell Biol. 55, 200-209, 1991  
 A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti  
 A:Reference number: A38872; MUID:92037656  
 A:Accession: A38872  
 A:Molecule type: mRNA  
 A:Residues: 1-877,968-1043 <KOC>  
 A:Cross-references: GB:S64268; GB:S64270  
 R:Gosdwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.  
 Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990  
 A:Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion  
 A:Reference number: A37785; MUID:91097553  
 A:Accession: A37785  
 A:Molecule type: mRNA  
 A:Residues: 44-123, 'V', 125-493 <GOO>  
 A:Cross-references: GB:IM58165; NID:9162966; PIDN:AAA62709.1; PID:9552318  
 R:Zimbelmann, R.  
 Submitted to the EMBL Data Library, February 1991  
 A:Reference number: S38721  
 A:Accession: S38721  
 A:Molecule type: mRNA  
 A:Residues: 44-1043 <ZIM>  
 A:Cross-references: EMBL:X57784; NID:9436061; PIDN:CAA40930.1; PID:9436062  
 R:Koch, P.J.; Walsh, M.T.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.  
 Eur. J. Cell Biol. 53, 1-12, 1990  
 A:Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m  
 A:Reference number: A48173; MUID:91168965  
 A:Accession: A48173  
 A:Molecule type: mRNA  
 A:Residues: 44-1001, 'AQPPSAR', <KOC>  
 A:Cross-references: GB:X57784  
 A:Note: this sequence has been revised in references A38872 and S38721  
 C:Genetics:  
 A:Gene: DSG1  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-49/Domain: propeptide #status predicted <PRO>  
 F:50-1043/Domain: extracellular #status predicted <EXT>  
 F:50-548/Domain: extracellular #status predicted <EXT>  
 F:52-157/Domain: cadherin repeat homology <CR1>  
 F:160-269/Domain: cadherin repeat homology <CR2>  
 F:372-385/Domain: cadherin repeat homology <CR3>  
 F:392-491/Domain: cadherin repeat homology <CR4>  
 F:549-574/Domain: transmembrane #status predicted <TM>  
 F:575-1043/Domain: intracellular #status predicted <INT>  
 F:846-875/Domain: desmoglein repeat <DG1>  
 F:876-905/Domain: desmoglein repeat <DG2>  
 F:906-933/Domain: desmoglein repeat <DG3>  
 F:934-962/Domain: desmoglein repeat <DG4>  
 F:963-1013/Domain: desmoglein repeat <DG5>  
 F:110/Binding site: glycyne/serine-rich  
 F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:180/496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 73.7%; Score 56; DB 1; Length 1043;  
 Best Local Similarity 60.0%; Pred. No. 0.037;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 TPEFLSRTGVEVT 15  
 Db 206 SPMTIRNTGEIRT 220  
 :||||:|||||

RESULT 4  
 S32892  
 hypothetical protein 6 (pefi 5' region) - Salmonella typhimurium  
 C:Species: Salmonella typhimurium  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Sep-1999  
 C:Accession: S32892  
 R:Friedrich, M.J.; Kinsey, N.E.; Villa, J.; Kadner, R.J.  
 Mol. Microbiol. 8, 543-558, 1993

A:Title: Nucleotide sequence of a 13.9kb segment of the 90kb virulence plasmid of Salmon  
 A:Reference number: S32886; MUID:93316852  
 A:Accession: S32892  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-295 <FRI>  
 A:Cross-references: EMBL:108613  
 C:Superfamily: Salmonella typhimurium hypothetical protein 6 (pefi 5' region)

Query Match 57.2%; Score 43.5; DB 2; Length 295;  
 Best Local Similarity 80.0%; Pred. No. 1.9;  
 Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1 TPFLLSRNTGTVRT 15  
 |||T:|T|  
 Db 132 TLMFSLSR-TGEVRT 145

RESULT 5  
 S38673  
 desmoglein 2 - human  
 N:Alternate names: desmoglein HDCC  
 C:Species: Homo sapiens (man)  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999  
 C:Accession: S38673; B38872  
 R:Zimbelmann, R.  
 Submitted to the EMBL Data Library, September 1993  
 A:Reference number: S38673  
 A:Accession: S38673  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1117 <ZIN>  
 A:Cross-references: EMBL:226317; NID:9416177; PID:CAA81226.1; PID:9416178  
 R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
 Eur. J. Cell Biol. 55, 200-208, 1991  
 A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide  
 A:Reference number: A38872; MUID:92037656  
 A:Accession: B38872  
 A:Molecule type: mRNA  
 A:Residues: 777-1117 <KOC>  
 A:Cross-references: GB:S64273  
 C:Genetics:  
 A:Gene: GDB:DSG2  
 A:Cross-references: GDB:128808; OMIM:125671  
 A:Map position: 18q12.1-18q12.2  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein  
 F:51-158/Domain: cadherin repeat homology <CRL>  
 F:161-271/Domain: cadherin repeat homology <CR2>

Query Match 56.6%; Score 43; DB 2; Length 1117;  
 Best Local Similarity 50.0%; Pred. No. 9.3;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 TPFLLSRNTGTVRT 15  
 |||T:|T|  
 Db 208 PVFYLKDTGEIYT 221

RESULT 6  
 YKMY  
 citrate (sl)-synthase (EC 4.1.3.7) - Mycobacterium smegmatis  
 C:Species: Mycobacterium smegmatis  
 C>Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 18-Jun-1999  
 C:Accession: S17168  
 R:David, M.; Lubinski-Mink, S.; Ben-Zvi, A.; Sussa, M.; Ullitzur, S.; Kuhn, J.  
 Biochem. J. 278, 225-234, 1991  
 A:Title: Citrate synthase from Mycobacterium smegmatis. Cloning, sequence determination  
 A:Reference number: S17168; MUID:91354207  
 A:Accession: S17168  
 A:Molecule type: DNA

A:Residues: 1-375 <DAY>  
 A:Cross-references: EMBL:X60513; NID:944513; PID:CAA43028.1; PID:944514  
 C:Genetics:  
 A:Gene: gltA  
 C:Superfamily: citrate (sl)-synthase  
 C:Keywords: acetyl-CoA; carbon-carbon lyase; oxo-acid-lyase; tricarboxylic acid cycle  
 F:227,266,317/Active site: His, His, Asp #status predicted

Query Match 52.6%; Score 40; DB 1; Length 375;  
 Best Local Similarity 63.6%; Pred. No. 10;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TPFLLSRNTG 11  
 |||T:|T|  
 Db 335 TPLFMSRTTG 345

RESULT 7  
 S52694  
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN1 - yeast (Saccharomyces  
 N:Alternate names: protein P9659.3; protein YPR145W  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 19-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 16-Jul-1999  
 C:Accession: S52694; S69033  
 R:Dang, V.D.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.  
 Submitted to the EMBL Data Library, March 1995  
 A:Description: Multiple regulatory systems control expression of the Saccharomyces ce  
 A:Reference number: S52694  
 A:Accession: S52694  
 A:Molecule type: DNA  
 A:Residues: 1-572 <DAN>  
 A:Cross-references: EMBL:248675; NID:91163061; PID:CAA88594.1; PID:9747902  
 R:Fallon, L.  
 Submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of S. cerevisiae cosmid 9659.  
 A:Reference number: S69019  
 A:Accession: S69033  
 A:Molecule type: DNA  
 A:Residues: 1-572 <FUL>  
 A:Cross-references: EMBL:U40829; NID:91066476; PID:AA86284.1; PID:91066479; MIPS:YF  
 C:Genetics:  
 A:Gene: SGD:ASN1  
 A:Cross-references: SGD:S0006349; MIPS:YPR145W  
 A:Map position: 16R  
 C:Superfamily: asparagine synthase (glutamine-hydrolyzing)  
 C:Keywords: asparagine biosynthesis; ligase  
 F:2/Active site: Cys (amide transfer) #status predicted

Query Match 52.6%; Score 40; DB 2; Length 572;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPFLLSR 8  
 |||T:|T|  
 Db 347 TPFLLSR 354

RESULT 8  
 S55982  
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) ASN2 - yeast (Saccharomyces  
 N:Alternate names: protein G6358; protein YCR124W  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 16-Jul-1999  
 C:Accession: S55982; S64433  
 R:van Dyck, L.; Goffeau, A.  
 Submitted to the EMBL Data Library, December 1994  
 A:Description: Genes for an asn synthase, a GlcF-motif nucleoporin and a putative hom  
 e new ORFs, remnants of Ty and three tRNA genes.  
 A:Reference number: S55976  
 A:Accession: S55982  
 A:Molecule type: DNA

A:Residues: 1-572 <VAM>  
 A:Cross-references: EMBL:X83099; NID:g642340; PIDN:CA56159.1; PID:g642347  
 R:Van Dyck, L.; Skala, J.; de Weiglisse, P.; Punelle, B.; Talla, E.; Nawrocki, A.; Del  
 submitted to the Protein Sequence Database, May 1996  
 A:Reference number: S64428  
 A:Accession: S64433  
 A:Molecule type: DNA  
 A:Residues: 1-572 <VAM>  
 A:Cross-references: EMBL:Z72909; NID:g1323202; PIDN:CA97135.1; PID:e243463; PID:g132320  
 A:Experimental source: strain S288C  
 A:Gene: SGD:ASN2  
 A:Cross-references: SGD:S0003356; MIPS:YGR124w  
 A:Map position: 7R  
 C:Superfamily: asparagine synthase (glutamine-hydrolyzing)  
 C:Keywords: asparagine biosynthesis; ligase  
 F:2/Active site: Cys (amide transfer) #status predicted

Query Match 52.6%; Score 40; DB 2; Length 572;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 Db 346 TPMFLSR 353

RESULT 9  
 F69294  
 DNA repair protein RAD25 homolog - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Sep-1999  
 C:Accession: F69294  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 ; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Usterbeck, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.  
 Smith, H.O.; Moese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343  
 A:Accession: F69294  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-452 <KLE>  
 A:Cross-references: GB:AE001080; GB:AE000782; NID:g2689403; PIDN:AB90879.1; PID:g265027  
 C:Superfamily: unassigned DEAD/H box helicases; DEAD/H box helicase homology  
 C:Keywords: ATP; P-loop  
 F:95-417/Domain: DEAD/H box helicase homology <DEAD>  
 F:95-102/Region: nucleotide-binding motif A (P-loop)  
 F:175-180/Region: nucleotide-binding motif B  
 F:175-182/Region: DEXH motif

Query Match 52.6%; Score 40; DB 2; Length 452;  
 Best Local Similarity 72.7%; Pred. No. 13;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 LLSRNTGEV 15  
 Db 430 LLSRNTGEV 440

RESULT 10  
 D70539  
 Probable citrate synthase - Mycobacterium tuberculosis (strain H37RV)  
 C:Species: Mycobacterium tuberculosis  
 C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Jun-1999  
 C:Accession: D70539  
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garler, T.; Churcher, C.; Harris, D.; Gordon, S.  
 Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998  
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno  
 A:Reference number: A70500; MUID:98295987  
 A:Accession: D70539  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-393 <COL>  
 A:Cross-references: GB:Z95585; GB:AL123456; NID:g3261787; PIDN:CA90942.1; PID:e31715  
 A:Experimental source: strain H37RV  
 C:Gene: glcA1  
 C:Superfamily: citrate (sl)-synthase

Query Match 51.3%; Score 39; DB 2; Length 393;  
 Best Local Similarity 63.6%; Pred. No. 17;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 TPMFLSRNTG 11  
 Db 350 TPMFLSRNTG 360

RESULT 11  
 A71536  
 Hypothetical protein CR260 - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C:Species: Chlamydia trachomatis  
 C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 21-Nov-1998  
 C:Accession: A71536  
 R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
 Science 282, 754-759, 1998  
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
 A:Reference number: A71570; MUID:9900809  
 A:Accession: A71536  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-163 <AN>  
 A:Cross-references: GB:AE001299; GB:AE001273; NID:g3328671; PID:g3328672  
 C:Experimental source: serotype D, strain UW-3/Cx  
 C:Gene: CR260

Query Match 51.3%; Score 39; DB 2; Length 163;  
 Best Local Similarity 58.3%; Pred. No. 6.6;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 PMFLSRNTGEV 13  
 Db 43 PMFLSRNTGEV 54

RESULT 12  
 C72756  
 Probable multiple sugar-binding transport system permease protein APE0040 - Aeropyrum  
 C:Species: Aeropyrum pernix  
 C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: C72756  
 R:Karabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.  
 DNA Res. 6, 83-101, 1999  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero  
 A:Reference number: A72450; MUID:99310339  
 A:Accession: C72756  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-302 <KAW>  
 A:Cross-references: DDBJ:AP000056; NID:g5103388; PIDN:BAA78949.1; PID:d1042725; PID:g  
 A:Experimental source: strain KI  
 C:Gene: APE0040

Query Match 51.3%; Score 39; DB 2; Length 302;  
 Best Local Similarity 50.0%; Pred. No. 13;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 TPWFLLSR 15  
 |||||:  
 DB 150 TPWFLLSR 163

RESULT 13  
 A:PMN2  
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 2 - garden pea  
 N:Alternate names: asparagine synthetase (glutamine-hydrolyzing)  
 C:Species: Pisum sativum (garden pea)  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
 C:Accession: S11443  
 R:Tsal, F.Y.; Coruzzi, G.M.  
 EMBL J. 9, 323-332, 1990  
 A>Title: Dark-induced and organ-specific expression of two asparagine synthetase genes  
 A:Reference number: S11443; MUID:90151804  
 A:Molecule type: DNA  
 A:Residues: 1-583 <RSA>  
 A:Cross-references: EMBL:X52180; NID:g20651; PIDN:CAA36430.1; PID:g20652  
 C:Genetics:  
 A:Gene: AS2  
 C:Superfamily: asparagine synthase (glutamine-hydrolyzing)  
 C:Keywords: asparagine biosynthesis; ligase  
 F:/Active site: Cys (amide transfer) #status predicted

Query Match 50.0%; Score 38; DB 1; Length 583;  
 Best Local Similarity 87.5%; Pred. No. 39;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPWFLLSR 8  
 |||||:  
 DB 322 TPWFLLSR 329

RESULT 14  
 A:PMN1  
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 - garden pea  
 N:Alternate names: asparagine synthetase (glutamine-hydrolyzing)  
 C:Species: Pisum sativum (garden pea)  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 18-Jun-1999  
 C:Accession: S11444  
 R:Tsal, F.Y.; Coruzzi, G.M.  
 EMBL J. 9, 323-332, 1990  
 A>Title: Dark-induced and organ-specific expression of two asparagine synthetase genes  
 A:Reference number: S11443; MUID:90151804  
 A:Molecule type: DNA  
 A:Residues: 1-586 <RSA>  
 A:Cross-references: EMBL:X52179; NID:g20649; PIDN:CAA36429.1; PID:g20650  
 C:Comment: This protein is one of a family of glutamine amidotransferases that have dual  
 : an amniator domain, which catalyzes the ammonia-dependent reaction, and a glutamine am  
 C:Genetics:  
 A:Gene: AS1  
 C:Superfamily: asparagine synthase (glutamine-hydrolyzing)  
 C:Keywords: asparagine biosynthesis; ligase  
 F:/Active site: Cys (amide transfer) #status predicted

Query Match 50.0%; Score 38; DB 1; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 39;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPWFLLSR 8  
 |||||:  
 DB 323 TPWFLLSR 330

RESULT 15  
 S69182  
 asparagine synthase (glutamine-hydrolyzing) (EC 6.3.5.4) 1 - Lotus japonicus

C:Species: Lotus japonicus  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 16-Jul-1999  
 C:Accession: S69182; S57931  
 R:Waterhouse, R.N.; Smyth, A.J.; Massonneau, A.; Prosser, I.M.; Clarkson, D.T.  
 Plant Mol. Biol. 30, 883-897, 1996  
 A>Title: Molecular cloning and characterisation of asparagine synthetase from Lotus  
 A:Reference number: S69182; MUID:96270368  
 A:Accession: S69182  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-586 <MAT>  
 A:Cross-references: EMBL:X89409; NID:g897770; PIDN:CAA61589.1; PID:g897771  
 A:Experimental source: strain B-129  
 C:Genetics:  
 A:Gene: AS1  
 C:Superfamily: asparagine synthase (glutamine-hydrolyzing)  
 C:Keywords: asparagine biosynthesis; ligase

Query Match 50.0%; Score 38; DB 2; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 39;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPWFLLSR 8  
 |||||:  
 DB 322 TPWFLLSR 329

Search completed: January 7, 2000, 08:52:42  
 Job time: 2272 sec

Sat Jan 15 11:45:01 2000

us-08-991-628-4.rpr

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:35 : Search time 28.55 seconds

(without alignments)  
15.282 Million cell updates/sec

Title: US-08-991-628-2

Perfect score: 78

Sequence: 1 FGIFVYDKNTGIDINI 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38:\*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	78	100.0	999	1	DSG3_HUMAN	P22926 homo sapien
2	56	71.8	1117	1	DSG2_HUMAN	Q14126 homo sapien
3	55	70.5	1043	1	DSG1_BOVIN	O03763 bos taurus
4	55	70.5	1049	1	DSG1_HUMAN	O02413 homo sapien
5	45	57.7	883	1	DSG1_BOVIN	O01107 bos taurus
6	45	57.7	884	1	DSG1_HUMAN	O08554 homo sapien
7	43	55.1	782	1	CAD5_PIG	O02840 sus scrofa
8	43	55.1	790	1	CAD6_HUMAN	P55285 homo sapien
9	43	55.1	789	1	CAD6_RAT	P55280 rattus norv
10	43	55.1	789	1	CAD6_HUMAN	P79955 gallus gall
11	43	55.1	790	1	CAD6_HUMAN	O13634 homo sapien
12	43	55.1	566	1	YGA3_YEAST	P33157 saccharomyc
13	42	53.8	713	1	CAD1_HUMAN	P53250 homo sapien
14	42	53.8	866	1	DSG1_MOUSE	P55849 mus musculu
15	41	52.6	794	1	CAD6_HUMAN	P55289 homo sapien
16	40	51.3	266	1	PRCB_YEAST	P02657 saccharomyc
17	40	51.3	1637	1	TP2B_CHICK	O42131 gallus gall
18	40	51.3	1612	1	TP2B_CRILLO	O04359 cricetus
19	40	51.3	1626	1	TP2B_HUMAN	O02880 homo sapien
20	40	51.3	1612	1	TP2B_MOUSE	O64511 mus musculu
21	39	50.0	1651	1	FTN2_HAEMIN	P43708 haemophilus
22	39	50.0	1073	1	ITAG_HUMAN	P33229 homo sapien
23	39	50.0	1073	1	ITAG_MOUSE	O61179 mus musculu
24	39	50.0	1356	1	KAB7_YEAST	P31374 saccharomyc
25	39	50.0	646	1	VEI_HPV07	O05133 human papil
26	39	50.0	1748	1	YNR2_YEAST	P33886 saccharomyc
27	38	48.7	887	1	CAD1_CHICK	P08641 gallus gall
28	38	48.7	790	1	CAD6_MOUSE	P73326 mus musculu
29	38	48.7	881	1	DPOI_HPBVS	P03161 ground squi
30	38	48.7	433	1	FTS2_ARATH	O42545 arabidopsis
31	38	48.7	273	1	HLPA_HAEMIN	P31728 haemophilus
32	38	48.7	456	1	NIFE_CLOPA	P10996 clostridium
33	38	48.7	626	1	PARC_BOBBO	O10666 bovine bu
34	38	48.7	88	1	RS15_MYCEL	P06789 mycoplasma
35	38	48.7	657	1	VEI_HPV18	P06857 human papil
36	38	48.7	94	1	YRHE_ECOLI	P46857 escherichia
37	37	47.4	921	1	AKH_DAUCA	P37142 daucus caro
38	37	47.4	882	1	CAD1_HUMAN	P12830 homo sapien
39	37	47.4	884	1	CAD1_MOUSE	P09803 mus musculu

## ALIGNMENTS

RESULT	1	STANDARD	PRT	999 AA
DSG3_HUMAN	1			
1	DSG3_HUMAN			
2	DSG3_HUMAN			
3	DSG3_HUMAN			
4	DSG3_HUMAN			
5	DSG3_HUMAN			
6	DSG3_HUMAN			
7	DSG3_HUMAN			
8	DSG3_HUMAN			
9	DSG3_HUMAN			
10	DSG3_HUMAN			
11	DSG3_HUMAN			
12	DSG3_HUMAN			
13	DSG3_HUMAN			
14	DSG3_HUMAN			
15	DSG3_HUMAN			
16	DSG3_HUMAN			
17	DSG3_HUMAN			
18	DSG3_HUMAN			
19	DSG3_HUMAN			
20	DSG3_HUMAN			
21	DSG3_HUMAN			
22	DSG3_HUMAN			
23	DSG3_HUMAN			
24	DSG3_HUMAN			
25	DSG3_HUMAN			
26	DSG3_HUMAN			
27	DSG3_HUMAN			
28	DSG3_HUMAN			
29	DSG3_HUMAN			
30	DSG3_HUMAN			
31	DSG3_HUMAN			
32	DSG3_HUMAN			
33	DSG3_HUMAN			
34	DSG3_HUMAN			
35	DSG3_HUMAN			
36	DSG3_HUMAN			
37	DSG3_HUMAN			
38	DSG3_HUMAN			
39	DSG3_HUMAN			
40	DSG3_HUMAN			
41	DSG3_HUMAN			
42	DSG3_HUMAN			
43	DSG3_HUMAN			
44	DSG3_HUMAN			
45	DSG3_HUMAN			
46	DSG3_HUMAN			
47	DSG3_HUMAN			
48	DSG3_HUMAN			
49	DSG3_HUMAN			
50	DSG3_HUMAN			
51	DSG3_HUMAN			
52	DSG3_HUMAN			
53	DSG3_HUMAN			
54	DSG3_HUMAN			
55	DSG3_HUMAN			
56	DSG3_HUMAN			
57	DSG3_HUMAN			
58	DSG3_HUMAN			
59	DSG3_HUMAN			
60	DSG3_HUMAN			
61	DSG3_HUMAN			
62	DSG3_HUMAN			
63	DSG3_HUMAN			
64	DSG3_HUMAN			
65	DSG3_HUMAN			
66	DSG3_HUMAN			
67	DSG3_HUMAN			
68	DSG3_HUMAN			
69	DSG3_HUMAN			
70	DSG3_HUMAN			
71	DSG3_HUMAN			
72	DSG3_HUMAN			
73	DSG3_HUMAN			
74	DSG3_HUMAN			
75	DSG3_HUMAN			
76	DSG3_HUMAN			
77	DSG3_HUMAN			
78	DSG3_HUMAN			
79	DSG3_HUMAN			
80	DSG3_HUMAN			
81	DSG3_HUMAN			
82	DSG3_HUMAN			
83	DSG3_HUMAN			
84	DSG3_HUMAN			
85	DSG3_HUMAN			
86	DSG3_HUMAN			
87	DSG3_HUMAN			
88	DSG3_HUMAN			
89	DSG3_HUMAN			
90	DSG3_HUMAN			
91	DSG3_HUMAN			
92	DSG3_HUMAN			
93	DSG3_HUMAN			
94	DSG3_HUMAN			
95	DSG3_HUMAN			
96	DSG3_HUMAN			
97	DSG3_HUMAN			
98	DSG3_HUMAN			
99	DSG3_HUMAN			
100	DSG3_HUMAN			
101	DSG3_HUMAN			
102	DSG3_HUMAN			
103	DSG3_HUMAN			
104	DSG3_HUMAN			
105	DSG3_HUMAN			
106	DSG3_HUMAN			
107	DSG3_HUMAN			
108	DSG3_HUMAN			
109	DSG3_HUMAN			
110	DSG3_HUMAN			
111	DSG3_HUMAN			
112	DSG3_HUMAN			
113	DSG3_HUMAN			
114	DSG3_HUMAN			
115	DSG3_HUMAN			
116	DSG3_HUMAN			
117	DSG3_HUMAN			
118	DSG3_HUMAN			
119	DSG3_HUMAN			
120	DSG3_HUMAN			
121	DSG3_HUMAN			
122	DSG3_HUMAN			
123	DSG3_HUMAN			
124	DSG3_HUMAN			
125	DSG3_HUMAN			
126	DSG3_HUMAN			
127	DSG3_HUMAN			
128	DSG3_HUMAN			
129	DSG3_HUMAN			
130	DSG3_HUMAN			
131	DSG3_HUMAN			
132	DSG3_HUMAN			
133	DSG3_HUMAN			
134	DSG3_HUMAN			
135	DSG3_HUMAN			
136	DSG3_HUMAN			
137	DSG3_HUMAN			
138	DSG3_HUMAN			
139	DSG3_HUMAN			
140	DSG3_HUMAN			
141	DSG3_HUMAN			
142	DSG3_HUMAN			
143	DSG3_HUMAN			
144	DSG3_HUMAN			
145	DSG3_HUMAN			
146	DSG3_HUMAN			
147	DSG3_HUMAN			
148	DSG3_HUMAN			
149	DSG3_HUMAN			
150	DSG3_HUMAN			
151	DSG3_HUMAN			
152	DSG3_HUMAN			
153	DSG3_HUMAN			
154	DSG3_HUMAN			
155	DSG3_HUMAN			
156	DSG3_HUMAN			
157	DSG3_HUMAN			
158	DSG3_HUMAN			
159	DSG3_HUMAN			
160	DSG3_HUMAN			
161	DSG3_HUMAN			
162	DSG3_HUMAN			
163	DSG3_HUMAN			
164	DSG3_HUMAN			
165	DSG3_HUMAN			
166	DSG3_HUMAN			
167	DSG3_HUMAN			
168	DSG3_HUMAN			
169	DSG3_HUMAN			
170	DSG3_HUMAN			
171	DSG3_HUMAN			
172	DSG3_HUMAN			
173	DSG3_HUMAN			
174	DSG3_HUMAN			
175	DSG3_HUMAN			
176	DSG3_HUMAN			
177	DSG3_HUMAN			
178	DSG3_HUMAN			
179	DSG3_HUMAN			
180	DSG3_HUMAN			
181	DSG3_HUMAN			
182	DSG3_HUMAN			
183	DSG3_HUMAN			
184	DSG3_HUMAN			
185	DSG3_HUMAN			
186	DSG3_HUMAN			
187	DSG3_HUMAN			
188	DSG3_HUMAN			
189	DSG3_HUMAN			
190	DSG3_HUMAN			
191	DSG3_HUMAN			
192	DSG3_HUMAN			
193	DSG3_HUMAN			
194	DSG3_HUMAN			
195	DSG3_HUMAN			
196	DSG3_HUMAN			
197	DSG3_HUMAN			
198	DSG3_HUMAN			
199	DSG3_HUMAN			
200	DSG3_HUMAN			
201	DSG3_HUMAN			
202	DSG3_HUMAN			
203	DSG3_HUMAN			
204	DSG3_HUMAN			
205	DSG3_HUMAN			
206	DSG3_HUMAN			
207	DSG3_HUMAN			
208	DSG3_HUMAN			
209	DSG3_HUMAN			
210	DSG3_HUMAN			
211	DSG3_HUMAN			
212	DSG3_HUMAN			
213	DSG3_HUMAN			
214	DSG3_HUMAN			
215	DSG3_HUMAN			
216	DSG3_HUMAN			
217	DSG3_HUMAN			
218	DSG3_HUMAN			
219	DSG3_HUMAN			
220	DSG3_HUMAN			
221	DSG3_HUMAN			
222	DSG3_HUMAN			
223	DSG3_HUMAN			
224	DSG3_HUMAN			
225	DSG3_HUMAN			
226	DSG3_HUMAN			
227	DSG3_HUMAN			
228	DSG3_HUMAN			
229	DSG3_HUMAN			
230	DSG3_HUMAN			
231	DSG3_HUMAN			
232	DSG3_HUMAN			
233	DSG3_HUMAN			
234	DSG3_HUMAN			
235	DSG3_HUMAN			
236	DSG3_HUMAN			
237	DSG3_HUMAN			
238	DSG3_HUMAN			
239	DSG3_HUMAN			
240	DSG3_HUMAN			
241	DSG3_HUMAN			
242	DSG3_HUMAN			
243	DSG3_HUMAN			
244	DSG3_HUMAN			
245	DSG3_HUMAN			
246	DSG3_HUMAN			
247	DSG3_HUMAN			
248	DSG3_HUMAN			
249	DSG3_HUMAN			
250	DSG3_HUMAN			
251	DSG3_HUMAN			
252	DSG3_HUMAN			
253	DSG3_HUMAN			
254	DSG3_HUMAN			
255	DSG3_HUMAN			
256	DSG3_HUMAN			
257	DSG3_HUMAN			
258	DSG3_HUMAN			
259	DSG3_HUMAN			
260	DSG3_HUMAN			
261	DSG3_HUMAN			
262	DSG3_HUMAN			
263	DSG3_HUMAN			
264	DSG3_HUMAN			
265	DSG3_HUMAN			
266	DSG3_HUMAN			
267	DSG3_HUMAN			
268	DSG3_HUMAN			
269	DSG3_HUMAN			
270	DSG3_HUMAN			
271	DSG3_HUMAN			
272	DSG3_HUMAN			
273	DSG3_HUMAN			
274	DSG3_HUMAN			
275	DSG3_HUMAN			
276	DSG3_HUMAN			
277	DSG3_HUMAN			
278	DSG3_HUMAN			
279	DSG3_HUMAN			
280	DSG3_HUMAN			
281	DSG3_HUMAN			
282	DSG3_HUMAN			
283	DSG3_HUMAN			
284	DSG3_HUMAN			
285	DSG3_HUMAN			
286	DSG3_HUMAN			
287	DSG3_HUMAN			
288	DSG3_HUMAN			
289	DSG3_HUMAN			
290	DSG3_HUMAN			
291	DSG3_HUMAN			
292	DSG3_HUMAN			
293	DSG3_HUMAN			
294	DSG3_HUMAN			
295	DSG3_HUMAN			
296	DSG3_HUMAN			
297	DSG3_HUMAN			
298	DSG3_HUMAN			
299	DSG3_HUMAN			
300	DSG3_HUMAN			
301	DSG3_HUMAN			
302	DSG3_HUMAN			
303	DSG3_HUMAN			
304	DSG3_HUMAN			
305	DSG3_HUMAN			
306	DSG3_HUMAN			
307	DSG3_HUMAN			
308	DSG3_HUMAN			
309	DSG3_HUMAN			
310	DSG3_HUMAN			
311	DSG3_HUMAN			
312	DSG3_HUMAN			
313	DSG3_HUMAN			
314	DSG3_HUMAN			
315	DSG3_HUMAN			
316	DSG3_HUMAN			
317	DSG3_HUMAN			
318	DSG3_HUMAN			
319	DSG3_HUMAN			
320	DSG3_HUMAN			
321	DSG3_HUMAN			
322	DSG3_HUMAN			
323	DSG3_HUMAN			
324	DSG3_HUMAN			
325	DSG3_HUMAN			
326	DSG3_HUMAN			
327	DSG3_HUMAN			
328	DSG3_HUMAN			
329	DSG3_HUMAN			
330	DSG3_HUMAN			
331	DSG3_HUMAN			
332	DSG3_HUMAN			
333	DSG3_HUMAN			
334	DSG3_HUMAN			
335	DSG3_HUMAN			
336	DSG3_HUMAN			
337	DSG3_HUMAN			
338	DSG3_HUMAN			
339	DSG3_HUMAN			
340	DSG3_HUMAN			
341	DSG3_HUMAN			
342	DSG3_HUMAN			
343	DSG3_HUMAN			
344	DSG3_HUMAN			
345	DSG3_HUMAN			
346	DSG3_HUMAN			
347	DSG3_HUMAN			
348	DSG3_HUMAN			
349	DSG3_HUMAN			
350	DSG3_HUMAN			
351	DSG3_HUMAN			
352	DSG3_HUMAN			
353	DSG3_HUMAN			
354	DSG3_HUMAN			
355	DSG3_HUMAN			
356	DSG3_HUMAN			
357	DSG3_HUMAN			
358	DSG3_HUMAN			
359	DSG3_HUMAN			
360	DSG3_HUMAN			
361	DSG3_HUMAN			
362	DSG3_HUMAN			
363	DSG3_HUMAN			
364	DSG3_HUMAN			
365	DSG3_HUMAN			
366				



FT CARBOHYD 180 180 POTENTIAL.  
 FT CARBOHYD 459 459 POTENTIAL.  
 FT CARBOHYD 545 545 POTENTIAL.  
 SQ SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;

Query Match 100.0%; Score 78; DB 1; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDIINI 15  
 Db 97 FGIFVVDKNTGDIINI 111

RESULT 2  
 DSG2\_HUMAN  
 ID DSG2\_HUMAN STANDARD; PRT; 1117 AA.  
 AC 014126;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 2 PRECURSOR (HDGC).  
 GN DSG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-COLON CARCINOMA;  
 RX MEDLINE: 94192736  
 RA SCHAEFER S., KOCH P.J., FRANK W.W.;  
 RT "Identification of the ubiquitous human desmoglein, Dsg2, and the  
 RT expression catalogue of the desmoglein subfamily of desmosomal  
 RT cadherins.";  
 RL Exp. Cell Res. 211:391-399(1994).  
 RN [2]  
 RP SEQUENCE OF 777-1117 FROM N.A.  
 RX MEDLINE: 92037656.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT polypeptide and identification of a second type of desmoglein gene.";  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL). BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z26317; CA81226.1; -  
 DR HSP: P15116; INCI.  
 DR MIM: 125671; -  
 DR PFM: PFO0028; cadherin; 4.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KM Cell adhesion: Glycoprotein; Transmembrane; Repeat; signal:  
 KW Cytoskeleton; Calcium-binding; POTENTIAL.  
 FT SIGNAL 1 23  
 FT PROPEP 24 48  
 FT CHAIN 49 1117  
 FT DOMAIN 49 608  
 FT TRANSMEM 609 633  
 FT DOMAIN 634 1117  
 CYTOPLASMIC (POTENTIAL).

FT REPEAT 49 159 CADHERIN 1.  
 FT REPEAT 160 272 CADHERIN 2.  
 FT REPEAT 273 387 CADHERIN 3.  
 FT REPEAT 388 502 CADHERIN 4.  
 FT REPEAT 880 911 DESMOGLEIN REPEAT 1.  
 FT REPEAT 912 941 DESMOGLEIN REPEAT 2.  
 FT REPEAT 942 967 DESMOGLEIN REPEAT 3.  
 FT REPEAT 968 991 DESMOGLEIN REPEAT 4.  
 FT REPEAT 992 1020 DESMOGLEIN REPEAT 5.  
 FT REPEAT 1021 1050 DESMOGLEIN REPEAT 6.  
 FT CARBOHYD 111 111 POTENTIAL.  
 FT CARBOHYD 181 181 POTENTIAL.  
 FT CARBOHYD 308 308 POTENTIAL.  
 FT CARBOHYD 461 461 POTENTIAL.  
 FT CARBOHYD 513 513 POTENTIAL.  
 SQ SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match 71.8%; Score 56; DB 1; Length 1117;  
 Best Local Similarity 60.0%; Pred. No. 0.13;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDIINI 15  
 Db 98 FGIFVVDKNTGDIINI 112

RESULT 3  
 DSG1\_BOVIN  
 ID DSG1\_BOVIN STANDARD; PRT; 1043 AA.  
 AC 003763;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).  
 GN DSG1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-MUZZLE EPITHELIUM;  
 RX MEDLINE: 91168965.  
 RA KOCH P.J., WALSH M.J., SCHWELZ M., GOLDSCHMIDT M.D.,  
 RA ZIMBELMANN R., FRANK W.W.;  
 RT "Identification of desmoglein, a constitutive desmosomal  
 RT glycoprotein, as a member of the cadherin family of cell adhesion  
 RT molecules.";  
 RL Eur. J. Cell Biol. 53:1-12(1990).  
 RN [3]  
 RP REVISIONS, AND SEQUENCE OF 101-123.  
 RX MEDLINE: 92037656.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
 RA FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT polypeptide and identification of a second type of desmoglein gene.";  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 RN [4]  
 RP SEQUENCE OF 44-493 FROM N.A.  
 RX MEDLINE: 91097553.  
 RA GOODMAN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COMIN P.;  
 RT "Desmoglein shows extensive homology to the cadherin family of cell  
 RT adhesion molecules.";  
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.

```

CC -1- SURCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, MZZLE, TONGUE AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X58466; CAA41380.1; -
DR EMBL: X57784; CAA40930.1; -
DR EMBL: M58165; AAA62709.1; -
DR PIR: S14603; IYB0G1.
DR HSSP: P09803; 1EDR.
DR PFAM: PF00028; cadherin; 3.
DR PROSITE: PS00232; CADHERIN; 2.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1043
FT TRANSMEM 50 548
FT DOMAIN 549 573
FT TRANSMEM 574 1043
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 498
FT REPEAT 499 845
FT REPEAT 846 875
FT REPEAT 876 905
FT REPEAT 906 933
FT REPEAT 934 962
FT DOMAIN 963 1012
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT CARBOHYD 496 496
FT CONFLICT 124 124
FT SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;

Query Match 70.5%; Score 55; DB 1; Length 1043;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 FGIFVVDKNTGDIINI 15
Db 97 YGIFVINOKTGEINI 111

```

```

RA BUXTON R.S., MAGER A.I.;
RT "Desmosomal glycoprotein Dg1, a component of intercellular desmosome
RT junctions, is related to the cadherin family of cell adhesion
RT molecules."
RL Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).
CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.
CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE
CC FILAMENTS MEDIATING CELL-CELL ADHESION.
CC -1- SURCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.
CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS
CC (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X56654; CAA39976.1; -
DR PIR: S16906; IYH0G1.
DR HSSP: P09803; 1EDR.
DR PFAM: PF00028; cadherin; 4.
DR PROSITE: PS00232; CADHERIN; 2.
KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
KW Calcium-binding; Repeat.
FT SIGNAL 1 23
FT PROPEP 24 49
FT CHAIN 50 1049
FT TRANSMEM 50 545
FT DOMAIN 546 570
FT TRANSMEM 571 1049
FT REPEAT 50 158
FT REPEAT 159 270
FT REPEAT 271 385
FT REPEAT 386 497
FT REPEAT 498 839
FT REPEAT 840 869
FT REPEAT 870 899
FT REPEAT 900 927
FT REPEAT 928 956
FT DOMAIN 969 1019
FT CARBOHYD 36 36
FT CARBOHYD 110 110
FT CARBOHYD 180 180
FT SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;

Query Match 70.5%; Score 55; DB 1; Length 1049;
Best Local Similarity 60.0%; Pred. No. 0.17;
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;
QY 1 FGIFVVDKNTGDIINI 15
Db 97 YGIFVINOKTGEINI 111

```

```

RESULT 4
ID DSG1_HUMAN STANDARD; PRT; 1049 AA.
AC 002413;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DG1).
GN DSG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KERATINOCYTES;
RX MEDLINE: 91371279.
RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALLOTTIS P., POYNTER D.,
RA ANNEWMANN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,

```

```

RESULT 5
ID DSG1_BOVIN STANDARD; PRT; 893 AA.
AC 001107; Q28095;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).
GN DSG1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

```

OC Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.  
 RC TISSUE-EPIDERMIS;  
 RX MEDLINE: 91185414.  
 RA COLLINS J.E., LEGAN P.K., KENNY T.P., MACGARVIE J., HOLTON J.L.,  
 GARROD D.R.;  
 RT "Cloning and sequence analysis of desmosomal glycoproteins 2 and 3  
 (desmocollins): cadherin-like desmosomal adhesion molecules with  
 heterogeneous cytoplasmic domains.";  
 RL J. Cell Biol. 113:381-391(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. (1B), AND PARTIAL SEQUENCE.  
 RX MEDLINE: 91239591.  
 RA MECHANIC S., RAYNOR K., HILL J.E., COMIN P.;  
 RT "Desmocollins form a distinct subset of the cadherin family of cell  
 adhesion molecules.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4476-4480(1991).  
 RN [3]  
 RP SEQUENCE OF 133-893 FROM N.A. (1A), AND PARTIAL SEQUENCE.  
 RC TISSUE-MUZZLE EPITHELIUM;  
 RX MEDLINE: 92008912.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., SCHMELZ M.,  
 FRANK W.W.;  
 RT "Amino acid sequence of bovine muzzle epithelial desmocollin derived  
 from cloned cDNA: a novel subtype of desmosomal cadherins.";  
 RL Differentiation 47:29-36(1991).  
 RN [4]  
 RP SEQUENCE OF 133-155.  
 RX MEDLINE: 91115997.  
 RA HOLTON J.L., KENNY T.P., LEGAN P.K., COLLINS J.E., KEEN J.N.,  
 SHARMA R., GARROD D.R.;  
 RT "Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal  
 similarity to calcium-dependent cell-cell adhesion molecules.";  
 RL J. Cell Sci. 97:239-246(1990).  
 RN [5]  
 RP PHOSPHORYLATION.  
 RX MEDLINE: 91009551.  
 RA PARRISH E.P., MARSTON J.E., MATTEY D.L., MEASURES H.R., VENNING R.,  
 GARROD D.R.;  
 RT "Site heterogeneity, phosphorylation and transmembrane organisation  
 of desmosomal glycoproteins 2 and 3 (desmocollins) in MDCK cells.";  
 RL J. Cell Sci. 96:239-248(1990).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED  
 IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS  
 MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL  
 POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL  
 ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED  
 TO THE KERATINIZATION OF EPITHELIAL TISSUES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS: 1A OR DG2 (SHOWN HERE) AND 1B OR  
 DG3. ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- TISSUE SPECIFICITY: EPIDERMIS AND WEAKLY IN TONGUE PAPILLAE.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 (POTENTIAL).  
 CC -1- PTM: FORM 1A IS PHOSPHORYLATED ON A SERINE BUT FORM 1B IS NOT.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOCOLLIN SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isp.slb.ch/announce/>  
 CC or send an email to [license@slb.slb.ch](mailto:license@slb.slb.ch)).  
 CC EMBL: M67489; AAA30492.1; -  
 DR EMBL: X56966; CAA40286.1; -  
 DR EMBL: X56967; CAA40287.1; -  
 DR EMBL: X56968; CAA40288.1; -  
 DR EMBL: X56969; CAA40289.1; -  
 DR EMBL: X58029; CAA41088.1; -  
 DR PIR: B38456; IJBDF.

DR PIR: A43838; IJBDF.  
 DR PIR: A39377; A39377.  
 DR HSSP: P09803; 1EDH.  
 DR PIR: P00028; cadherin: 5.  
 DR PROSITE: PS00232; CADHERIN: 2.  
 KW Cell adhesion; glycoprotein; Phosphorylation; Transmembrane; Signal;  
 KM Repeat; Cytoskeleton; Calcium-binding; Alternative splicing.  
 FT SIGNAL  
 FT PROPEP 30 132  
 FT CHAIN 133 893  
 FT DOMAIN 133 692  
 FT TRANSMEM 693 715  
 FT DOMAIN 716 893  
 FT REPEAT 133 240  
 FT REPEAT 241 352  
 FT REPEAT 353 470  
 FT REPEAT 471 574  
 FT REPEAT 575 682  
 FT CARBOHYD 163 163  
 FT CARBOHYD 398 398  
 FT CARBOHYD 545 545  
 FT VARIANT 519 519  
 FT VARIANT 788 788  
 FT VARSPLIC 829 839  
 FT VARSPLIC 840 893  
 FT CONFLICT 485 485  
 FT SEQUENCE 893 AA; 99647 MW; F55D78B CRC32;  
 SO  
 Query Match 57.7%; Score 45; DB 1; Length 893;  
 Best Local Similarity 53.8%; Pred. No. 6.8;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 Oy 1 RGIFVVDKNTGDI 13  
 Db 180 FNLFFIERKDTGDI 192  
 RESULT 6  
 DSCI\_HUMAN STANDARD; PRT; 894 AA.  
 ID DSCI\_HUMAN  
 AC Q08554;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DESMOCOLLIN 1A/1B PRECURSOR (DESMOSOMAL GLYCOPROTEIN 2/3) (DG2/DG3).  
 GN DSCI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FORESKIN;  
 RX MEDLINE: 93283249.  
 RA THEIS D.G., KOCH P.J., FRANK W.W.;  
 RT "Differential synthesis of type 1 and type 2 desmocollin mRNAs in  
 human stratified epithelia.";  
 RL Int. J. Dev. Biol. 37:101-110(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-FORESKIN;  
 RX ZIMBELMANN R.;  
 RT Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-SKIN;  
 RX MEDLINE: 94116981.  
 RA KING I.A., ARNEBANN J., SPURR N.K., BUXTON R.S.;  
 RT "Cloning of the cDNA (DSCI) coding for human type 1 desmocollin and  
 RL its assignment to chromosome 18.";  
 RL Genomics 18:185-194(1993).  
 RN [4]

RP SEQUENCE OF 135-151 AND 283-292.  
 RX MEDLINE: 91323543.  
 RA KING I.A., MAGEE A.I., REES D.A., BUXTON R.S.;  
 RT "Keratinization is associated with the expression of a new protein  
 related to the desmosomal cadherins Dbl/111.";  
 RL FEBS Lett. 286:9-12(1991).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED  
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS  
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL  
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL  
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED  
 CC TO THE KERATINIZATION OF EPITHELIAL TISSUES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS; 1A OR DG2 (SHOWN HERE) AND 1B OR  
 CC DG3; ARE PRODUCED BY ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- TISSUE SPECIFICITY: STRONGLY EXPRESSED IN EPIDERMIS, LESS IN LYMPH  
 CC NODE AND TONGUE.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: 234532; CAA84279.1;  
 CC EMBL: 234532; CAA84278.1;  
 CC EMBL: X72925; CAA51429.1;  
 CC DR EMBL: X72925; CAA51429.1;  
 CC DR HSSP: P09803; 1EDH.  
 CC DR MIM: 125643;  
 CC DR PFAM: PF00028; cadherin; 5.  
 CC DR PROSITE: PS00332; CADHERIN; 3.  
 CC KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;  
 CC KW Cytoskeleton; Calcium-binding; Alternative splicing.  
 CC FT SIGNAL 1 29  
 CC FT PROPEP 30 134  
 CC FT CHAIN 135 894  
 CC FT DOMAIN 135 691  
 CC FT TRANSMEM 692 714  
 CC FT DOMAIN 715 894  
 CC FT REPEAT 242  
 CC FT REPEAT 243  
 CC FT REPEAT 355  
 CC FT REPEAT 471  
 CC FT REPEAT 575  
 CC FT REPEAT 576  
 CC FT REPEAT 682  
 CC FT CARBOHYD 165 155  
 CC FT CARBOHYD 546 546  
 CC FT VARSPLIC 830 840  
 CC FT VARSPLIC 841 894  
 CC FT CONFLICT 132 132  
 CC FT SEQUENCE 894 AA; 100044 MW; A287BCA9 CRC32;  
 SQ  
 Query Match 57.7%; Score 45; DB 1; Length 894;  
 Best Local Similarity 53.8%; Pred. No. 6.8;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).  
 GN CDH5.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Suidae; Sus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA KILSHAW P.J.;  
 RL Submitted (Jun-1997) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A  
 CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE  
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT  
 CC ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON  
 CC (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL  
 CC BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: Y13919; CAA74225.1;  
 CC HSSP: P09803; 1EDH.  
 CC DR PFAM: PF00028; cadherin; 5.  
 CC DR PFAM: PF01049; Cadherin, C-term; 1.  
 CC DR PROSITE: PS00332; CADHERIN; 3.  
 CC KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 CC KW Calcium-binding; Repeat; Signal.  
 CC FT SIGNAL 1 22  
 CC FT PROPEP 23 44  
 CC FT CHAIN 45 782  
 CC FT DOMAIN 45 592  
 CC FT TRANSMEM 593 619  
 CC FT DOMAIN 620 782  
 CC FT REPEAT 148  
 CC FT REPEAT 149  
 CC FT REPEAT 256  
 CC FT REPEAT 370  
 CC FT REPEAT 371  
 CC FT REPEAT 475  
 CC FT REPEAT 476  
 CC FT REPEAT 592  
 CC FT DOMAIN 736 751  
 CC FT CARBOHYD 58 58  
 CC FT CARBOHYD 154 154  
 CC FT CARBOHYD 360 360  
 CC FT CARBOHYD 440 440  
 CC FT CARBOHYD 532 532  
 CC FT CARBOHYD 534 534  
 CC FT SEQUENCE 782 AA; 87546 MW; 007F70E0 CRC32;  
 SQ  
 Query Match 55.1%; Score 43; DB 1; Length 782;  
 Best Local Similarity 63.6%; Pred. No. 13;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDI 13  
 Db 182 FNFVYIKDGTGDI 194  
 RESULT 7  
 ID CAD5\_PIG  
 AC 002840;  
 DT 15-JUL-1998 (Rel. 36, Created)

OY 3 IFVVDKNTGDI 13  
 Db 89 VERVDENTGDV 99  
 RESULT 8  
 ID CAD6\_HUMAN  
 AC P55285;  
 DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE CADHERIN-6 PRECURSOR (KIDNEY-CADHERIN) (K-CADHERIN).  
 GN CDH6.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95262134.  
 RA SHIMOYAMA Y., GOTOH M., TERASAKI T., KITAJIMA M., HIROHASHI S.;  
 RT "Isolation and sequence analysis of human cadherin-6 complementary  
 RT DNA for the full coding sequence and its expression in human  
 RT carcinoma cells."  
 RL Cancer Res. 55:2206-2211(1995).  
 RN [2]  
 RP SEQUENCE OF 377-790 FROM N.A.  
 RC TISSUE-FETAL BRAIN;  
 RX MEDLINE: 91283540.  
 RA SUZUKI S., SANO K., TANIHARA H.;  
 RT "Diversity of the cadherin family: evidence for eight new cadherins  
 RT in nervous tissue."  
 RL Cell Regul. 2:261-270(1991).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN, CEREBELLUM, AND  
 CC KIDNEY, LUNG, PANCREAS, AND GASTRIC MUCOSA SHOW A WEAK EXPRESSION.  
 CC ALSO EXPRESSED IN CERTAIN LIVER AND KIDNEY CARCINOMAS.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: D31784; BAA0562.1; -  
 DR HSSP: P15116; INCI.  
 DR MIM: 603007; -  
 DR PFAM: PF00028; cadherin; 5.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KW Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 53  
 FT CHAIN 54 790  
 FT DOMAIN 54 615  
 FT TRANSMEM 616 636  
 FT DOMAIN 637 790  
 FT REPEAT 34 139  
 FT REPEAT 160 268  
 FT REPEAT 269 383  
 FT REPEAT 384 486  
 FT REPEAT 487 608  
 FT CARBOHYD 49 49  
 FT CARBOHYD 255 255  
 FT CARBOHYD 399 399  
 FT CARBOHYD 437 437  
 FT CARBOHYD 455 455  
 FT CARBOHYD 536 536  
 FT CONFLICT 421 421  
 FT CONFLICT 425 425  
 FT SEQUENCE 790 AA; 88308 MW; FC415ELE CRC32;

Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 QY 3 IFVVDKNTGDI 13  
 Db 100 LFLINNTGDI 110  
 RESULT 9  
 ID CAD6\_RAT STANDARD; PRT: 789 AA.  
 AC P55280;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DE CADHERIN-6 PRECURSOR (KIDNEY-CADHERIN) (K-CADHERIN).  
 GN CDH6 OR KCAD.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ACI; TISSUE-KIDNEY;  
 RX MEDLINE: 94243827.  
 RA XIANG Y.Y., TANAKA M., SUZUKI M., IGARASHI H., KIYOKAWA E., NAITO Y.,  
 RA OHTAWARA Y., SHEN Q., SUGIMURA H., KINO I.;  
 RT "Isolation of complementary DNA encoding K-cadherin, a novel rat  
 RT cadherin preferentially expressed in fetal kidney and kidney  
 RT carcinoma."  
 RL Cancer Res. 54:3034-3041(1994).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN KIDNEY AND BRAIN.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
 CC -----  
 DR EMBL: D25290; BAA04975.1; -  
 DR HSSP: P15116; INCI.  
 DR PFAM: PF00028; cadherin; 5.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KW Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 18  
 FT PROPEP 19 53  
 FT CHAIN 54 789  
 FT DOMAIN 54 615  
 FT TRANSMEM 616 636  
 FT DOMAIN 637 789  
 FT REPEAT 34 139  
 FT REPEAT 160 268  
 FT REPEAT 269 383  
 FT REPEAT 384 486  
 FT REPEAT 487 608  
 FT CARBOHYD 49 49  
 FT CARBOHYD 255 255  
 FT CARBOHYD 399 399  
 FT CARBOHYD 437 437  
 FT CARBOHYD 455 455  
 FT CARBOHYD 536 536  
 FT SEQUENCE 789 AA; 88340 MW; 403CCCB3 CRC32;

Query Match 55.1%; Score 43; DB 1; Length 790;  
 Best Local Similarity 54.5%; Pred. No. 13;

Query Match 55.1%; Score 43; DB 1; Length 789;  
 Best Local Similarity 54.5%; Pred. No. 13;

Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 13  
: : : : :  
Db 100 LFIINTGDI 110

RESULT 10  
ID CADA\_CHICK STANDARD; PRT; 789 AA.  
AC P79995;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CADERIN-10 PRECURSOR.  
GN CDH10.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;  
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN, AND RETINA;  
RX FUSHIMI D., ARDIT K., TAKEICHI M., REDIES C.;  
RA "Cloning and expression analysis of cadherin-10 in the CNS of the  
RT chicken embryo."  
RL Dev. Dyn. 209:269-285(1997).  
CC -1- FUNCTION: CADERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC MANNER IN CONNECTING CELLS: CADERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CADERIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AB000512; BAA19130.1; -  
DR HSSP: P09803; 1EDH.  
DR PFAM: PF00028; cadherin; 5.  
DR PFAM: PF01049; Cadherin\_C-term; 1.  
DR PROSITE: PS00232; CADERIN; 3.  
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
KW Calcium-binding; Repeat; Signal.  
FT SIGNAL 1 22  
FT PROPEP 23 54  
FT CHAIN 55 789  
FT DOMAIN 55 606  
FT TRANSMEM 607 634  
FT TRANSMEM 607 634  
FT DOMAIN 635 789  
FT REPEAT 55 160  
FT REPEAT 161 269  
FT REPEAT 270 384  
FT REPEAT 385 487  
FT REPEAT 488 606  
FT CARBOHYD 256 256  
FT CARBOHYD 438 438  
FT CARBOHYD 456 456  
FT CARBOHYD 534 534  
SO SEQUENCE 789 AA; 88392 MW; D44464F0 CRC32;

Query Match 55.1%; Score 43; DB 1; Length 789;  
Best Local Similarity 50.0%; Pred. No. 13;  
Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 14  
: : : : :  
Db 100 IFVVDKNTGDI 14

RESULT 12

Db 101 LFIIDKNTGDI 112

RESULT 11  
ID CADA\_HUMAN STANDARD; PRT; 790 AA.  
AC 013634;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE CADERIN-14 PRECURSOR.  
GN CDH14.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BRAIN;  
RX SHIBATA T., SHIMOYAMA Y., GOTOH M., HIROHASHI S.;  
RA "Identification of human cadherin-14, a novel neurally specific type  
RT II cadherin, by protein interaction cloning."  
RL J. Biol. Chem. 272:5236-5240(1997).  
CC -1- FUNCTION: CADERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
CC MANNER IN CONNECTING CELLS: CADERINS MAY THUS CONTRIBUTE TO THE  
CC SORTING OF HETEROGENEOUS CELL TYPES.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE CADERIN FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U59325; AAB02933.1; -  
DR HSSP: P15116; INC1.  
DR PFAM: PF00028; cadherin; 5.  
DR PFAM: PF01049; Cadherin\_C-term; 1.  
DR PROSITE: PS00232; CADERIN; 3.  
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
KW Calcium-binding; Repeat; Signal.  
FT SIGNAL 1 24  
FT PROPEP 25 53  
FT CHAIN 54 790  
FT DOMAIN 54 608  
FT TRANSMEM 609 636  
FT TRANSMEM 609 636  
FT DOMAIN 637 790  
FT REPEAT 34 159  
FT REPEAT 160 268  
FT REPEAT 269 383  
FT REPEAT 384 486  
FT REPEAT 487 608  
FT CARBOHYD 36 36  
FT CARBOHYD 255 255  
FT CARBOHYD 455 455  
FT CARBOHYD 536 536  
SO SEQUENCE 790 AA; 88072 MW; 4596258F CRC32;

Query Match 55.1%; Score 43; DB 1; Length 790;  
Best Local Similarity 58.3%; Pred. No. 13;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 14  
: : : : :  
Db 100 IFIIDDITGDI 111

RESULT 12

YGA3\_YEAST STANDARD; PRT; 566 AA.  
 ID YGA3\_YEAST  
 AC P5197;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHELICAL 62.8 KD TRP-ASP REPEATS CONTAINING PROTEIN IN PMCI-TRG2  
 DE INTERGENIC REGION.  
 GN YGL003C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomycetes.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA HEBLING U., HOFMANN B., DELIUS H.;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; Z7525; CAA96703.1; -  
 CC DR PFAM; PF00400; WD40; 4.  
 CC DR PROSITE; PS00678; WD\_REPEATS; 1.  
 KM Hypothetical protein: Repeat: WD repeat.  
 FT DOMAIN 32 38  
 FT REPEAT 300 330 WD1.  
 FT REPEAT 342 370 WD2.  
 FT REPEAT 383 413 WD3.  
 FT REPEAT 425 458 WD4.  
 FT REPEAT 469 501 WD5.  
 FT REPEAT 513 543 WD6.  
 FT SEQUENCE 566 AA; 62821 MW; 65B12728 CRC32;  
 SQ  
 OY 3 IFVVDKNTGDI 13  
 II: I I I I I I I  
 Db 285 IFLTDNNTGDV 295  
 Query Match 55.1%; Score 43; DB 1; Length 566;  
 Best Local Similarity 63.6%; Pred. No. 9;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 RESULT 13  
 CADD\_HUMAN STANDARD; PRT; 713 AA.  
 ID CADD\_HUMAN  
 AC P53290;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE T-CADHERIN PRECURSOR (TRUNCATED-CADHERIN) (CADHERIN-13) (H-CADHERIN)  
 DE (HEART-CADHERIN).  
 GN CDH13 OR CDHH.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TANIHARA H., SANO K., HEIMARK R.L., ST JOHN T., SUZUKI S.;  
 RL MEDLINE; 95073006.  
 RT Cloning of five human cadherins clarifies characteristic features of  
 RT cadherin extracellular domain and provides further evidence for two  
 RT structurally different types of cadherin.  
 RL Cell Adhes. Commun. 2:15-26(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RX MEDLINE; 96266329.  
 RA LEE S.W.;  
 RT "H-cadherin, a novel cadherin with growth inhibitory functions and  
 RT diminished expression in human breast cancer."  
 RL Nat. Med. 2:776-782(1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA SANO K., MORI Y., SANURADA A., FUJIMURA S., HORII A.;  
 RL Hum. Genet. 103:96-101(1998).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (BY  
 CC SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; L34058; AAA35624.1; -  
 CC DR EMBL; U59289; AAB18912.1; -  
 CC DR EMBL; U59288; AAB18911.1; -  
 CC DR EMBL; AB001103; BAA32411.1; -  
 CC DR EMBL; AB001090; BAA32411.1; JOINED.  
 CC DR EMBL; AB001091; BAA32411.1; JOINED.  
 CC DR EMBL; AB001092; BAA32411.1; JOINED.  
 CC DR EMBL; AB001093; BAA32411.1; JOINED.  
 CC DR EMBL; AB001094; BAA32411.1; JOINED.  
 CC DR EMBL; AB001095; BAA32411.1; JOINED.  
 CC DR EMBL; AB001096; BAA32411.1; JOINED.  
 CC DR EMBL; AB001097; BAA32411.1; JOINED.  
 CC DR EMBL; AB001098; BAA32411.1; JOINED.  
 CC DR EMBL; AB001099; BAA32411.1; JOINED.  
 CC DR EMBL; AB001100; BAA32411.1; JOINED.  
 CC DR EMBL; AB001101; BAA32411.1; JOINED.  
 CC DR EMBL; AB001102; BAA32411.1; JOINED.  
 CC DR HSSP; P15116; INCT.  
 CC DR MIM; 601364; -  
 CC DR PFAM; PF00028; cadherin; 5.  
 CC DR PROSITE; PS00232; CADHERIN; 3.  
 KM Cell adhesion; Glycoprotein; Phosphorylation; Calcium-Binding; Repeat;  
 KW GPI-anchor; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 138  
 FT CHAIN 139 693  
 FT PROPEP 694 713  
 FT REPEAT 139 245  
 FT REPEAT 246 363  
 FT REPEAT 364 477  
 FT REPEAT 478 583  
 FT REPEAT 583 693  
 FT CARBOHYD 52 52  
 FT CARBOHYD 86 86  
 FT CARBOHYD 382 382  
 FT CARBOHYD 500 500  
 FT CARBOHYD 530 530  
 FT CARBOHYD 598 598  
 FT CARBOHYD 638 638  
 FT CARBOHYD 671 671  
 FT LIPID 693 693  
 FT SEQUENCE 713 AA; 78286 MW; 8BC381E5 CRC32;  
 SQ  
 Query Match 53.8%; Score 42; DB 1; Length 713;  
 Best Local Similarity 42.9%; Pred. No. 17;

Matches	65	Conservative	6	Mismatches	2	Indels	0	Gaps	0
QY	2	GIFVADKNTGDINI 15							
		:::    :::							
Db	186	GIFRINENTGVSVS 199							
RESULT	14								
DSCL_MOUSE		STANDARD;	PRT;	886 AA.					
AC	P55849								
DT	01-NOV-1997	(Rel. 35, Created)							
DT	01-NOV-1997	(Rel. 35, Last sequence update)							
DT	15-JUL-1998	(Rel. 36, Last annotation update)							
DE	DESMOCOLLIN 1A/1B	PRECURSOR.							
GN	DSCL.								
OS	Mus musculus (Mouse).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;								
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.								
RC	[1]								
RP	SEQUENCE FROM N.A.								
RX	STRAIN-CS7BL/6; TISSUE-SKIN;								
RX	MEDLINE; 96420658.								
RA	KING I.A., O'BRIEN T.J., BUXTON R.S.;								
RT	"Expression of the 'skin-type' desmosomal cadherin Dscl is closely								
RT	linked to the keratinization of epithelial tissues during mouse								
RT	development."								
RL	J. Invest. Dermatol. 107:531-538(1996).								
CC	-1- FUNCTION: COMPONENT OF INTRACELLULAR DESMOSOME JUNCTIONS. INVOLVED								
CC	IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS								
CC	MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL								
CC	POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL								
CC	ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS. LINKED								
CC	TO THE KERATINIZATION OF EPITHELIAL TISSUES.								
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.								
CC	-1- ALTERNATIVE PRODUCTS: FORMS 1A AND 1B ARE PRODUCED BY ALTERNATIVE								
CC	SPLICING OF THE SAME GENE.								
CC	-1- DOMAIN: CALCINUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS								
CC	(POTENTIAL).								
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOSOMAL SUBFAMILY.								
CC	-----								
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration								
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation								
CC	at the European Bioinformatics Institute. There are no restrictions on								
CC	use by non-profit institutions as long as its content is in no way								
CC	modified and this statement is not removed. Usage by and for commercial								
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>								
CC	or send an email to <a href="mailto:license@sdb.ch">license@sdb.ch</a> ).								
CC	-----								
DR	EMBL; X97986; CAA66628.1; -								
DR	EMBL; X97986; CAA66629.1; -								
DR	HSSP; P09803; 1EDH.								
DR	MED; MG1:109173; DSCL.								
DR	PFAM; PF00028; cadherin.5.								
DR	PROSITE; PS00232; CADHERIN.2.								
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;								
KW	Calcium-binding; Repeat; Alternative splicing.								
FT	SIGNAL	1	29	POTENTIAL.					
FT	PROPEP	30	134	POTENTIAL.					
FT	CHAIN	135	886	DESMOCOLLIN 1A/1B.					
FT	DOMAIN	135	691	INTRACELLULAR (POTENTIAL).					
FT	TRANSSEM	692	714	POTENTIAL.					
FT	DOMAIN	715	886	CYTOPLASMIC (POTENTIAL).					
FT	REPEAT	135	242	CADHERIN 1.					
FT	REPEAT	243	354	CADHERIN 2.					
FT	REPEAT	355	471	CADHERIN 3.					
FT	REPEAT	472	575	CADHERIN 4.					
FT	REPEAT	576	682	CADHERIN 5.					
FT	CARBOHYD	130	130						

```

FT VASAPPLIC 840 886 MISSING (IN FORM 1B).
SQ SEQUENCE 886 AA; 98953 MM; 6C3C94BA CRC32;
OY 1 FGIFVVDKNTGDI 13 53.88; Score 42; DB 1; Length 886;
: : : : : Pred. No. 21;
Db 182 YNLFYERDGTGDI 194 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 15
CADC_HUMAN STANDARD; PRT; 794 AA.
AC P55289;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE BRAIN-CADHERIN PRECURSOR (BR-CADHERIN) (CADHERIN-12) (N-CADHERIN 2)
DE (CADHERIN, NEURAL TYPE, 2).
GN CDH12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN (1)
RP TISSUE-BRAIN;
RC MEDLINE: 95073006.
RX TANIHARA H., SANO K., HEIMARK R.L., ST JOHN T., SUZUKI S.;
"Cloning of five human cadherins clarifies characteristic features of
cadherin extracellular domain and provides further evidence for two
structurally different types of cadherin.";
RL Cell Adhes. Commun. 2:15-26(1994).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN CORTEX;
RX MEDLINE: 95249541.
RX SELIG S., BRUNO S., SCHARF J.M., WANG C.H., VITALE E., GILLIAM T.C.,
KUNKEL L.M.;
"Expressed cadherin pseudogenes are localized to the critical region
of the spinal muscular atrophy gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3702-3706(1995).
CC -1- FUNCTION: CADHERINS ARE CALCIIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L34057; AAA35623.1;
DR EMBL, L33477; ABA48539.1;
DR HSSP; P09803; 1EDH.
DR MIM; 600562;
DR PFAM; PF00028; cadherin; 5.
DR PFAM; PF01049; cadherin.C term; 1.
DR PROSITE; PS00232; CADHERIN; 2
DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat; Signal.
FT SIGNAL 1-23
FT PROPEP 24-54
FT CHAIN 55-794
FT DOMAIN 55-609
EXTRACELLULAR (POTENTIAL).

```



FT	TRANSMEM	610	637	POTENTIAL.
FT	DOMAIN	638	794	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	55	160	CADHERIN 1.
FT	REPEAT	161	269	CADHERIN 2.
FT	REPEAT	270	384	CADHERIN 3.
FT	REPEAT	385	487	CADHERIN 4.
FT	REPEAT	488	609	CADHERIN 5.
FT	CARBOHYD	256	256	POTENTIAL.
FT	CARBOHYD	456	456	POTENTIAL.
FT	CARBOHYD	537	537	POTENTIAL.
FT	CARBOHYD	545	545	POTENTIAL.
FT	CONFLICT	68	68	V -> M (IN REF. 2).
FT	CONFLICT	349	349	E -> D (IN REF. 2).
FT	CONFLICT	416	416	G -> S (IN REF. 2).
FT	CONFLICT	644	644	H -> D (IN REF. 2).
FT	CONFLICT	733	733	I -> Y (IN REF. 2).
FT	CONFLICT	761	761	A -> T (IN REF. 2).
SO	SEQUENCE	794 AA;	88274 MM;	78BF3F7E CRC32;

Query Match 52.68; Score 41; DB 1; Length 794;  
Best Local Similarity 50.08; Pred. No. 28;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 IFVVDKNTGDIN 14  
Db 101 VFTIDETTGDIH 112

Search completed: January 7, 2000, 13:25:37  
Job time: 389 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:18 ; Search time 59.21 Seconds  
(without alignments)  
15.518 Million cell updates/sec

Title: US-08-991-628-2  
Perfect score: 78  
Sequence: 1 FGIFVVDKNTGNDINI 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL\_11.\*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	67	85.9	993 11 035902	035902 mus musculu
2	51	65.4	785 13 090763	090763 gallus gall
3	43	55.1	214 2 055701	055701 synechocyst
4	43	55.1	1329 5 076356	076356 caenorhabdi
5	43	55.1	950 13 093508	093508 brachydanio
6	43	55.1	790 13 090762	090762 gallus gall
7	42	53.8	1069 4 060245	060245 homo sapien
8	42	53.8	1072 4 060246	060246 homo sapien
9	42	53.8	1200 4 060247	060247 homo sapien
10	42	53.8	3380 5 024292	024292 drosophila
11	41.5	53.2	148 12 037171	037171 heliothis a
12	41	52.6	4307 5 019319	019319 caenorhabdi
13	40	51.3	598 2 029907	029907 archaeoglob
14	40	51.3	1033 2 033373	033373 neisseria g
15	40	51.3	813 11 063315	063315 rattus norv
16	40	51.3	694 11 063561	063561 rattus norv
17	40	51.3	1035 13 057537	057537 xenopus lae
18	40	50.0	803 4 075281	075281 homo sapien
19	39	50.0	803 4 075283	075283 homo sapien
20	39	50.0	816 4 075284	075284 homo sapien
21	39	50.0	391 5 019859	019859 caenorhabdi
22	39	50.0	205 5 027742	027742 plasmodium
23	39	50.0	801 11 0920M3	0920M3 mus musculu
24	39	50.0	73 13 006275	006275 xenopus lae
25	39	50.0	790 13 091838	091838 xenopus lae

26	39	50.0	1087 13 091778	091778 xenopus lae
27	38.5	49.4	798 2 050288	050288 mycoplasma
28	38	48.7	245 1 058723	058723 methanococ
29	38	48.7	1055 2 095632	095632 rickettsia
30	38	48.7	1058 2 095661	095661 rickettsia
31	38	48.7	828 2 072602	072602 synechocyst
32	38	48.7	497 2 074864	074864 salmonella
33	38	48.7	341 2 086281	086281 lactococcus
34	38	48.7	914 4 095151	095151 homo sapien
35	38	48.7	831 5 091388	091388 caenorhabdi
36	38	48.7	423 10 065875	065875 pisum sativ
37	38	48.7	797 11 063418	063418 rattus norv
38	38	48.7	274 12 084181	084181 human papil
39	38	48.7	1616 12 09YKD6	09YKD6 tomato mosa
40	37.5	48.1	444 2 092905	092905 bacillus sp
41	37	47.4	359 1 027477	027477 methanobact
42	37	47.4	1060 2 095585	095585 rickettsia
43	37	47.4	1060 2 095659	095659 rickettsia
44	37	47.4	977 2 025308	025308 helicobacte
45	37	47.4	1616 12 041352	041352 tobacco mos

## ALIGNMENTS

RESULT 1  
ID 035902 PRELIMINARY; PRT; 993 AA.  
AC 035902;  
DT 01-JAN-1998 (TREMREL. 05, Created)  
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)  
DE 01-MAY-1999 (TREMREL. 10, Last annotation update)  
DE DESMOGLEIN 3 (FRAGMENT).  
GN DSG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RA ISHIKAWA H., LI K., Uitto J.;  
RL SUBMITTED (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; U86016; AAB65091.1;  
DR PFM; P00028; cadherin; 4.  
DR PROSITE; P50023; CADHERIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON\_TER  
FT SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 85.9%; Score 67; DB 11; Length 993;  
Best Local Similarity 86.7%; Pred. No. 0.0046;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 97 FGIFVVDKNTGNDINI 111  
OY 1 FGIFVVDKNTGNDINI 15  
DB 97 FGIFVVDKNTGNDINI 111  
RESULT 2  
ID 090763 PRELIMINARY; PRT; 785 AA.  
AC 090763;  
DT 01-NOV-1996 (TREMREL. 01, Created)  
DT 01-NOV-1996 (TREMREL. 01, Last sequence update)  
DT 01-MAY-1999 (TREMREL. 10, Last annotation update)  
DE CHICKEN CADHERIN-7.  
OS Gallus gallus (Chicken).  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
RN [1]  
RP SEQUENCE FROM N.A.

OS *Brachydanio rerio* (Zebrafish) (Zebra danio).  
 OS *Eukaryotes*; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii.  
 OC Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Cyprinidae; Rasbora; Danio.  
 [1]

RP SEQUENCE FROM N.A.  
 RA YAMAMOTO A., DE ROBERTIS E.M.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AF042191; AAC62385.1; -  
 DR PFM: PF00028; cadherin.5;  
 DR PROSITE: PS00232; CADHERIN.5;  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 950 AA; 105381 MW; 96B102F0 CRC32;

Query Match 55.1%; Score 43; DB 13; Length 950;  
 Best Local Similarity 63.6%; Pred. No. 45;  
 Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Db 615 LFSIDKNTGDI 625

RESULT 6  
 ID 090762 PRELIMINARY; PRT; 790 AA.  
 AC 090762;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE CHICKEN CADHERIN-6B.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;  
 OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WHITE LECHORN; TISSUE-BRAIN;  
 RX MEDLINE: 95309115.  
 RA NAKAGAWA S., TAKEICHI M.;  
 RT "Neural crest cell-cell adhesion controlled by sequential and  
 RT subpopulation-specific expression of novel cadherins.";  
 RL Development 121:1321-1332(1995).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: DA2149; BAA07720.1; -  
 DR PFM: PF00028; cadherin.5;  
 DR PFM: PF01049; Cadherin\_C-term. 1.  
 DR PROSITE: PS00232; CADHERIN.3;  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 790 AA; 88659 MW; DE01FC76 CRC32;

Query Match 55.1%; Score 43; DB 13; Length 790;  
 Best Local Similarity 54.5%; Pred. No. 37;  
 Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;  
 Db 100 LFINENTGDI 110

RESULT 7  
 ID 060245 PRELIMINARY; PRT; 1069 AA.  
 AC 060245;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PCDH7 (BH-PCDH)A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;  
 RL Genomics 0:0-0(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL: AB006755; BAA25194.1; -  
 DR PFM: PF00028; cadherin.6;  
 DR PROSITE: PS00232; CADHERIN.6;  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 1069 AA; 116104 MW; F1732B30 CRC32;

Query Match 53.8%; Score 42; DB 4; Length 1069;  
 Best Local Similarity 58.3%; Pred. No. 75;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 580 GIFAIDPDSGDI 591

RESULT 8  
 ID 060246 PRELIMINARY; PRT; 1072 AA.  
 AC 060246;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PCDH7 (BH-PCDH)B.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;  
 RL Genomics 0:0-0(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AB006756; BAA25195.1; -  
 DR PFM: PF00028; cadherin.6;  
 DR PROSITE: PS00232; CADHERIN.6;  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 1072 AA; 116463 MW; A3DF367C CRC32;

Query Match 53.8%; Score 42; DB 4; Length 1072;  
 Best Local Similarity 58.3%; Pred. No. 75;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Db 580 GIFAIDPDSGDI 591

RESULT 9  
 ID 060247 PRELIMINARY; PRT; 1200 AA.  
 AC 060247;  
 DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PCDH7 (BH-PCDH)C.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA YOSHIDA K., YOSHITOMO-NAKAGAWA K., SEKI N., SASAKI M., SUGANO S.;  
 RL Genomics 0:0-0(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AB006757; BAA25196.1; -  
 DR PFM: PF00028; cadherin.5;  
 DR PROSITE: PS00232; CADHERIN.5;  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 1200 AA; 130337 MW; 56F1CD33 CRC32;

Query Match 53.8%; Score 42; DB 4; Length 1200;  
 Best Local Similarity 58.3%; Pred. No. 85;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GIVVVDKNTGDI 13  
 Db 533 GIFAIDPSGDI 544

RESULT 10

Q24292 Q24292 PRELIMINARY; PRT; 3380 AA.  
 AC Q24292;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE ADHERIN  
 GN DACHSUS.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95324813.  
 RA CLARK H.F., BRENTUP D., SCHNEITZ K., BIEBER A., GOODMAN C., NOLL M.;  
 RT "Dachsous encodes a member of the cadherin superfamily that controls  
 imaginal disc morphogenesis in Drosophila."  
 RL Genes Dev. 9:1530-1542(1995)  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; L08811; AAA79329.1; -  
 DR FLYBASE; FBgn0000497; ds.  
 DR PFAM; PF00028; cadherin; 25.  
 DR PROSITE; PS00232; CADHERIN; 18.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 3380 AA; 366356 MW; 7D83FAC3 CRC32;

Query Match 53.8%; Score 42; DB 5; Length 3380;  
 Best Local Similarity 58.3%; Pred. No. 2.7e+02;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GIVVVDKNTGDI 13  
 Db 713 GIFRIDRSTGEI 724

RESULT 11

Q37171 Q37171 PRELIMINARY; PRT; 148 AA.  
 AC Q37171;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last annotation update)  
 DE 17K ORF.  
 OS Heliothis armigera entomopoxvirus (HaEPV).  
 OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Entomopoxvirinae.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA SRISKANTHA A., OSBORNE R.J., DALL D.J.;  
 RL J. Gen. Virol. 0:0-0(1997).  
 DR EMBL; AF017791; AAB96622.1; -  
 SQ SEQUENCE 148 AA; 17488 MW; A945DB09 CRC32;

Query Match 53.2%; Score 41.5; DB 12; Length 148;  
 Best Local Similarity 42.1%; Pred. No. 10;  
 Matches 8; Conservative 4; Mismatches 2; Indels 5; Gaps 1;

Qy 1 FGIF-----VVDKNTGDI 14  
 Db 17 YGLFGDFKLIIDKNTGYN 35

RESULT 12

Q19319 Q19319 PRELIMINARY; PRT; 4307 AA.  
 AC Q19319; Q19785; Q21606;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)  
 DE F25F2.2 PROTEIN.  
 GN F25F2.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA AINSOUGH R.;  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; Z35662; CAA84721.1; -  
 DR EMBL; Z35599; CAA84721.1; JOINED.  
 DR EMBL; Z34802; CAA84721.1; JOINED.  
 DR EMBL; Z35599; CAA84661.1; -  
 DR EMBL; Z35662; CAA84661.1; JOINED.  
 DR EMBL; Z34802; CAA84661.1; JOINED.  
 DR EMBL; Z34802; CAA84339.1; -  
 DR EMBL; Z35599; CAA84339.1; JOINED.  
 DR EMBL; Z35662; CAA84339.1; JOINED.  
 DR PFAM; PF00028; cadherin; 6.  
 DR PROSITE; PS00232; CADHERIN; 9.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 4307 AA; 478958 MW; 3486FC46 CRC32;

Query Match 52.6%; Score 41; DB 5; Length 4307;  
 Best Local Similarity 72.7%; Pred. No. 5.1e+02;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 IFVVDKNTGDI 13  
 Db 3467 LFRVDPNTGDI 3477

RESULT 13

Q29907 Q29907 PRELIMINARY; PRT; 598 AA.  
 AC Q29907;  
 DT 01-JAN-1998 (TRENBLrel. 05, Created)  
 DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
 DT 01-AUG-1999 (TRENBLrel. 11, Last annotation update)  
 DE PROBABLE TUNGSTEN-CONTAINING ALDEHYDE FERREDOXIN OXIDOREDUCTASE 3  
 DE (EC 1.2.7.-) (AOR-3).  
 GN AF0340.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE; 98049343.  
 RA KLENK H.P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,  
 RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,  
 RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRIDES N.C.,  
 RA FEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,  
 RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,  
 RA PETERSON S., REICH C.I., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,  
 RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T.,  
 RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M.,  
 RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,  
 RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,  
 RA VENTER J.C.;  
 RT "The complete genome sequence of the hyperthermophilic,  
 sulphate-reducing archaeon Archaeoglobus fulgidus."  
 RL Nature 390:364-370(1997).  
 CC -1- PATHWAY: PYROGLYCOLYTIC PATHWAY.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

CC -1- COFACTOR: EACH SUBUNIT CONTAINS A 4FE4S CLUSTER AND A TUNGSTEN  
 CC COFACTOR  
 CC (CONSISTING OF 2 MOLYBDOPTERIN MOLECULES COORDINATED TO A  
 CC TUNGSTEN BY A TOTAL OF 4 SULFUR LIGANDS) (BY SIMILARITY).  
 CC -1- DOMAIN: CAN BE DIVIDED INTO 3 DOMAINS.  
 CC DOMAIN I WHICH FORMS A BASE ON WHICH THE SADDLE-LIKE TUNGSTEN  
 CC PTERIN COFACTOR SITS AND, CONTAINS THE MAGNESIUM BINDING SITES.  
 CC DOMAINS II AND III SURROUND THE OPPOSITE SURFACE OF THE TUNGSTEN  
 CC COFACTOR AND ARE INVOLVED IN INTERACTIONS WITH THE DIFFERENT  
 CC METAL CENTERS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE AOR/FOR/GAPOR/CAR/HVOR FAMILY.  
 DR EMBL: AE001081; AAB90895.1; -  
 DR TIGR: AF0340; -  
 DR PFAM: PF01314; AFOR: 1.  
 KW Hypothetical protein; Oxidoreductase; Iron-sulfur; 4Fe-4S;  
 KW Multigene family; Tungsten.  
 FT DOMAIN 1 216 DOMAIN I.  
 FT DOMAIN 2 217 413 DOMAIN II.  
 FT DOMAIN 3 291 598 DOMAIN III.  
 FT METAL 291 291 IRON-SULFUR (BY SIMILARITY).  
 FT METAL 294 294 IRON-SULFUR (BY SIMILARITY).  
 FT METAL 298 298 IRON-SULFUR (BY SIMILARITY).  
 FT METAL 493 493 IRON-SULFUR (BY SIMILARITY).  
 SQ SEQUENCE 598 AA; 65666 MW; 909BAB0E CRC32;

Query Match 51.3%; Score 40; DB 1; Length 598;  
 Best Local Similarity 40.0%; Pred. No. 86;  
 Matches 6; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDIINI 15  
 |:::|::|::|  
 DB 355 FAMIITEKGVGDPII 369

RESULT 14  
 03373 PRELIMINARY; PRT: 1033 AA.  
 AC 03373;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)  
 DE PILC PROTEIN.  
 CN PILC.  
 OS Neisseria gonorrhoeae.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-GC-653;  
 RX MEDLINE: 98129088.  
 RA BAECKMAN M., KAEHLSTROM H., JONSSON A.B.;  
 RT "The phase-variable plus-associated protein PILC is commonly  
 RT expressed in clinical isolates of Neisseria gonorrhoeae, and shows  
 RT sequence variability among strains."  
 RL Microbiology 144:149-156(1998).  
 DR EMBL: AJ001121; CAA04547.1; -  
 SQ SEQUENCE 1033 AA; 112936 MW; B0593053 CRC32;

Query Match 51.3%; Score 40; DB 2; Length 1033;  
 Best Local Similarity 46.7%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDIINI 15  
 |::|::|::|::|  
 DB 804 YGIFDDDKGTGVKV 818

RESULT 15  
 063315 PRELIMINARY; PRT: 813 AA.  
 AC 063315;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE LONG TYPE PB-CADHERIN.  
 OS Rattus norvegicus (Rat).  
 OC Euarystota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Euarystota; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-WISTAR; TISSUE-BRAIN;  
 RX MEDLINE: 96212232.  
 RA SUGIMOTO K., HONDA S., YAMAMOTO T., UKI T., MONDEN M., KAI A.,  
 RA MASUMOTO K., NAKAMURA T.;  
 RT "Molecular cloning and characterization of a newly identified member  
 RT of the cadherin family, pb-cadherin."  
 RT J. Biol. Chem. 271:11548-11556(1996).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: D83348; BA11894.1; -  
 DR PFAM: PF00028; Cadherin; 5.  
 DR PFAM: PF01049; Cadherin\_C-term; 1.  
 DR PROSITE: PS00232; CADHERIN; 2.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 813 AA; 87978 MW; 6E4D9FID CRC32;

Query Match 51.3%; Score 40; DB 11; Length 813;  
 Best Local Similarity 58.3%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDIINI 14  
 |::|::|::|::|  
 DB 106 IFIDELTGDIIH 117

Search completed: January 12, 2000, 23:15:19  
 Job time: 182 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 6, 2000, 15:40:36 ; Search time 51.33 Seconds

(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-4  
Sequence: 1 TPMPFLSRNTEVRT 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36:\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76	100.0	999	R30742	Human pemphigus vu
2	76	100.0	614	W07908	Pemphigus vulgaris
3	76	100.0	15	W04844	Self epitope of de
4	76	100.0	15	W64816	Desmoglein 3 206-2
5	76	100.0	15	W78815	Desmoglein 3 prote
6	64	84.2	778	W15489	Pemphigus foliaceu
7	48	63.2	616	R86867	Rat protocadherin
8	43	56.6	560	W13009	Segment of desmoso
9	43	56.6	263	W13010	Segment of desmoso
10	39	51.3	249	W38513	S. pneumoniae poss
11	39	51.3	835	W80689	S. pneumoniae cati
12	38	50.0	586	R07998	Asparagine synthet
13	38	50.0	583	R07999	Asparagine synthet
14	37	48.7	407	P94879	ORF Incorporated w
15	37	48.7	218	W62750	Streptococcus pneu
16	36.5	48.0	387	R49162	Fowlpox virus prot
17	36.5	48.0	387	R58887	Fowlpox virus (FPV
18	36.5	48.0	387	W10726	Homology vector 44
19	36	47.4	878	R55060	Sequence of human
20	36	47.4	878	R55487	Human E-cadherin p
21	36	47.4	484	R99735	Retinoid X recepto
22	36	47.4	451	R99739	Retinoid X recepto
23	36	47.4	469	W03448	Retinoid-activate
24	36	47.4	191	W04532	Reverse transcriptase
25	35	46.1	609	P30206	Sequence encoded b
26	35	46.1	219	R38092	nodB. Isolated nod
27	35	46.1	1148	R43671	M. leprae rpoB gene
28	35	46.1	284	W27703	E. coli AldA-I pro
29	35	46.1	472	W40072	Human retinoid rec
30	35	46.1	220	W40251	Human wild-type E-
31	35	46.1	555	W84154	Human desaturase e
32	35	46.1	608	W84155	Human desaturase e
33	35	46.1	746	W84156	Human desaturase e
34	35	46.1	746	W85135	A desaturase enzym
35	35	46.1	555	W85133	A desaturase enzym
36	35	46.1	608	W85134	A desaturase enzym
37	34.5	45.4	163	W85134	S. pneumoniae aspa
38	34	44.7	610	P93578	Extracellular neut
39	34	44.7	609	R21414	Vibriolysin, a prot

40	34	44.7	792	1	R59765	Extracellular neut
41	34	44.7	609	1	W86187	Vibriolysin, usefu
42	33.5	44.1	720	1	W19266	Lactobacillus amy
43	33	43.4	478	1	R04881	Recombinant elasta
44	33	43.4	337	1	R12995	GAP-DH. Promoter o
45	33	43.4	30	1	R93960	Peptide combining

## ALIGNMENTS

RESULT	1	
ID	R30742	
AC	R30742 standard; Protein; 999 AA.	
DT	14-JUN-1993 (first entry)	
DE	Human pemphigus vulgaris 130KD antigen.	
KM	Pemphigus vulgaris; skin disease; autoantibodies;	
KW	keratinocyte cell surface antigen; glycoprotein; cell adhesion.	
OS	Homo sapiens.	
PN	US/798918-A.	
PD	15-DEC-1992.	
PF	27-NOV-1991; 798918.	
PR	27-NOV-1991; US-798918.	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICE.	
PI	Amagai M, Klaus-Kovtun V, Stanley JR;	
DR	WPI; 93-067436/08.	
DR	N-PSDB; Q35992.	
PT	DNA encoding pemphigus vulgaris antigen - useful in proteins for	
PS	diagnostic and therapeutic uses	
CC	Disclosure; Fig 7; 50pp; English.	
CC	This sequence is the pemphigus vulgaris 130KD antigen. The protein	
CC	and its encoding DNA may be used in the diagnosis and treatment of	
CC	pemphigus vulgaris. It is thought that the antigen may be a cell	
CC	adhesion molecule.	
SO	Sequence 999 AA;	

Query Match 100.0%; Score 76; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TPMPFLSRNTEVRT 15  
Db 206 TPMPFLSRNTEVRT 220

RESULT 2  
W07908  
ID W07908 standard; protein; 614 AA.  
AC W07908;  
DT 29-JAN-1997 (first entry)  
DE Pemphigus vulgaris antigen protein extracellular region.  
KM Autoantibody; immunoglobulin G; IgG1, fusion protein; diagnosis;  
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;  
KW dermatology.  
OS Homo sapiens.  
PN J08188540-A.  
PD 23-JUL-1996.  
PF 30-JUN-1995; 165632.  
PR 30-JUN-1994; JP-173291.  
PA (NISHI) NISHIKAWA T.  
DR WPI; 96-388562/39.  
PT Fused protein recognised by pemphigus vulgaris autoantibody -  
PT useful to treat and diagnose pemphigus vulgaris  
PS Claim 1; Page 7-9; 9pp; Japanese.  
CC W07908 represents the human pemphigus vulgaris (PV) antigen  
CC extracellular region. The PV antigen is produced in patients with  
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare  
CC (vesicles) of the skin and mucous membranes, which is fatal if  
CC untreated. The PV antigen was fused to a human IgG1 hinge region  
CC and the resulting fusion protein is useful to treat or diagnose



CC pemphigus vulgaris.  
SQ Sequence 614 AA;

Query Match 100.0%; Score 76; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 6.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLSRTNGEVRT 15  
|||||  
DB 205 TPEFLSRTNGEVRT 219

RESULT 3  
W04844  
ID W04844 standard; peptide: 15 AA.

AC 18-FEB-1997 (first entry)  
DE Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerantigen; self-epitope; antigen; autoimmune disease;  
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphomannomutase;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW Influenza; haemagglutinin; reovirus; sigma protein.  
OS Homo sapiens.  
PN W09627387-A1.  
PD 12-SEP-1996.  
PE 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PI (HARD) HARVARD COLLEGE.  
PA Strominger JL, Wucherpfennig KW;  
DR WPI: 96-425218/42.  
PT pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for  
PT identification of other self and non-self antigens implicated in  
PT autoimmune disease  
PS Claim 1; Page 40; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise  
CC either an isolated human non-collagen or non-mysin basic protein  
CC (MBP) polypeptide which is capable of tolerising an individual to an  
CC autoantigen; or an isolated human pathogen polypeptide capable of  
CC tolerising an individual to that polypeptide. In both cases, the  
CC polypeptide (whether self or non-self) includes an amino acid  
CC sequence corresponding to a sequence motif for a MHC class II  
CC protein, such as HLA-DR, which is associated with a human autoimmune  
CC disease and which binds to the polypeptide to activate autoreactive  
CC T-cells in individuals with the autoimmune disease. This peptide is  
CC derived from the human desmoglein 3 protein (amino acids 206-220)  
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
CC derived from the human desmoglein protein are described in W04841-47.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLSRTNGEVRT 15  
|||||  
DB 1 TPEFLSRTNGEVRT 15

RESULT 4  
W64816  
ID W64816 standard; peptide: 15 AA.  
AC W64816:  
DE Desmoglein-3 206-220.  
DT 29-SEP-1998 (first entry)  
KW Desmoglein; Dg; gene therapy; pemphigus vulgaris; microparticle;  
KW autoantigen; autoimmune disease; MHC.  
OS Homo sapiens.  
PN US5783567-A.

PD 21-JUL-1998.  
PF 22-JAN-1997; 787547.  
PR 22-JAN-1997; US-787547.  
PI (PANG-) PANGAEA PHARM INC.  
PA Curley JM, Hedley ML, Langer RS;  
DR WPI: 98-427077/36.  
PT Microparticle encapsulated nucleic acids - for recombinant  
PT expression of proteins e.g. in gene therapy

PS Disclosure; Column 4; 42pp; English.  
CC The patent describes a new preparation of microparticles each  
CC comprising a polymeric matrix and a nucleic acid. The polymeric  
CC matrix consists of one or more synthetic polymers having a solubility  
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);  
CC and at least 90% of the microparticles have a diameter of less than  
CC 100 microns. The microparticles are useful for the delivery of nucleic  
CC acids to phagocytic cells. In one embodiment the microparticles are  
CC less than 20 microns in diameter and the nucleic acid (preferably in  
CC closed circular form) includes an expression control sequence  
CC operatively linked to a coding sequence, where the expression product  
CC of the coding sequence is a polypeptide having a length and a sequence  
CC which permits it to bind to an MHC class I or II molecule. The  
CC expression product is thus an effective stimulator of an immune  
CC response in mammals. The present sequence, an antigenic portion of  
CC desmoglein 3, is an example of an MHC class II peptide which can be  
CC expressed by the nucleic acid. It is associated with pemphigus  
CC vulgaris.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLSRTNGEVRT 15  
|||||  
DB 1 TPEFLSRTNGEVRT 15

RESULT 5  
W78815  
ID W78815 standard; peptide: 15 AA.  
AC W78815:  
DT 17-NOV-1998 (first entry)

DE Desmoglein 3 protein fragment 206-220.  
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;  
KW class II associated peptide; pathogen; gene therapy; genetic disease;  
KW infection; downregulation; immune response.  
OS Homo sapiens.  
PS Synthetic.  
PN W09831398-A1.  
PD 23-JUL-1998.  
PE 22-JAN-1998; U01499.  
PR 06-JAN-1998; US-003253.  
PR 22-JAN-1997; US-787547.  
PI (PANG-) PANGAEA PHARM INC.  
PA Curley JM, Hedley ML, Langer RS, Lunsford LB;  
DR WPI: 98-427356/36.  
PT New preparations of microparticles - comprising a synthetic polymer  
PT matrix and nucleic acid comprising an expression vector for use in  
PT gene therapy  
PS Disclosure; Page 8; 101pp; English.  
CC A microparticle preparation (MP) has been developed, consisting of  
CC microparticles having a diameter of less than 100 mu m. The MP comprises:  
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers  
CC having a solubility in water of less than 1 mg/l; and (b) an expression  
CC vector selected from RNA molecules (at least 50% of which are closed  
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).  
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)  
CC a PM; and (b) a NAM comprising an expression control sequence operatively  
CC linked to a coding sequence, where the coding sequence encodes an  
CC expression product selected from: (i) a polypeptide at least 7 amino  
CC acids in length, having a sequence identical to the sequence of: (1) a  
CC fragment of a naturally-occurring mammalian protein; or (ii) a fragment

CC of a naturally-occurring protein from an infectious agent which infects  
 CC a mammal; (2) a peptide having a length and sequence which permits it to  
 CC bind to an MHC class I or II molecule; and (3) the polypeptide or the  
 CC peptide linked to a trafficking sequence. W69763 to W69765, and W78793  
 CC to W78897 are peptide fragments for use in the present invention. The  
 CC MPs are highly effective vehicles for the delivery of polynucleotides  
 CC into phagocytic cells. They can be used for gene therapy, e.g. for  
 CC treating genetic diseases, infections or tumours or for downregulating  
 CC an immune response.  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 76; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLSRNTGEVRT 15  
 Db 1 TPEFLSRNTGEVRT 15

RESULT 6  
 W15489  
 ID W15489 standard; Protein: 778 AA.

AC W15489;  
 DT 17-JUN-1997 (first entry)  
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.  
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
 KW extracellular region; antigen; hinge portion; skin;  
 KW dermatitis herpetiformis; fusion protein; detection; ss.  
 OS Chimeric - Homo sapiens.  
 FH Key  
 FT domain  
 FT 1.545  
 Location/Qualifiers

/note= "Pemphigus foliaceus antigen protein"

PN 309077800-A.  
 PD 25-MAR-1997.  
 PF 12-SEP-1995; 260899.  
 PR 12-SEP-1995; JP-260899.  
 PA (NISH/) NISHIKAWA T.  
 DR WPI: 97-241758/22.  
 DR P-PSDB: T66428.  
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked  
 PT through the hinge region used to treat pemphigus foliaceus  
 PS Claim 1, Page 10-12; 17pp; Japanese.  
 CC This sequence represents a fused protein recognised by pemphigus  
 CC foliaceus patient autoantibody which comprises the constant region  
 CC of IgG linked to the extracellular region of pemphigus foliaceus  
 CC antigen protein through the hinge portion. Pemphigus foliaceus is  
 CC a chronic, generalised, vesicular and scaling skin eruption similar  
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion  
 CC protein is useful to treat pemphigus foliaceus. The antigen is  
 CC especially administered through an adsorbent upon which the fusion  
 CC protein is immobilised via a carrier. The fusion protein is also  
 CC useful for detecting pemphigus foliaceus antibodies which is useful  
 CC in immunodiagnosis. The fusion protein has little or no side effects.

SQ Sequence 778 AA;

Query Match 84.2%; Score 64; DB 1; Length 778;  
 Best Local Similarity 66.7%; Pred. No. 0.0012;  
 Matches 10; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPEFLSRNTGEVRT 15  
 Db 206 SPMEIINRNGEIRT 220

RESULT 7  
 R86867  
 ID R86867 standard; Protein: 616 AA.  
 AC R86867;  
 DT 27-AUG-1996 (first entry)  
 DE Rat protocadherin pc5.

KW Protocadherin: pc3; pc4; pc5; human; rat; cadherin; cell adhesion;  
 KW catenin; therapy.

OS Rattus rattus.  
 PN W09600289-A1.

PD 04-JAN-1996.

PF 26-JUN-1995; U08071.

PR 27-JUN-1994; US-268161.  
 PA (DOHE-) DOHERTY EYE INST.

PI Suzuki S;  
 DR WPI: 96-068873/07.

DR N-PSDB: T03574.

PT polynucleotide(s) encoding human protocadherins pc3 and pc4 and rat  
 PT pc5 - involved in cell-cell adhesion and regulation activities

PS Claim 17; Page 130-132; 146pp; English.

CC R86865-R86867 represent the sequences for three protocadherins. This  
 CC sequence represents the rat protocadherin pc5. These sequences are

CC related to cadherin, and possess cell adhesive ability. Cadherins are  
 CC glycosylated integral membrane proteins that are involved in cell-cell

CC adhesion. Cadherins are composed of an N-terminal extracellular domain  
 CC which consists of 5 unique subdomains, a membrane spanning domain, and a

CC C-terminal cytoplasmic domain. The cytoplasmic domain interacts with the  
 CC cytoskeleton through catenins and other cytoskeleton associated proteins.

CC The cytoplasmic domain is not present in all cadherins, but in those  
 CC which possess it, it is essential for the cadherin adhesive function.

CC The cadherins which do not possess a cytoplasmic domain appear to  
 CC function via a different method from those with a cytoplasmic domain.

CC These sequences were isolated using primers 1 and 2 (see T03575 and  
 CC T03576). The proteins may have regulatory functions in the cell, as well

CC as the cell-cell adhesive properties. Antibodies produced against these  
 CC sequences are useful for modulating the binding activity of these

CC protocadherins, and can be used therapeutically.  
 SQ Sequence 616 AA;

Query Match 63.2%; Score 48; DB 1; Length 616;  
 Best Local Similarity 61.5%; Pred. No. 0.7;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPEFLSRNTGEV 13  
 Db 85 TPLFOLNMTGEI 97

RESULT 8

W13009  
 ID W13009 standard; Protein: 560 AA.

AC W13009;

DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.

KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;

KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.

OS Homo sapiens.  
 PN DE19531033-A1.

PD 27-FEB-1997.

PF 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.

PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;

DR WPI: 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on  
 PT surface of epithelial or carcinoma cells; not bound to desmosomes,  
 PT useful for diagnosis and treatment of carcinoma micrometastases

PS Claim 7; Page 5; 8pp; German.

CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or

CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to

CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separate, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to

CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or  
 CC carcinomas.  
 SO Sequence 560 AA;

Query Match 56.6%; Score 43; DB 1; Length 560;  
 Best Local Similarity 50.0%; Pred. No. 5.1;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 PMLLSRNTGEVRT 15  
 Db 160 PVEFLNKDTGEIYT 173

RESULT 9  
 W13010  
 ID W13010 standard; protein; 263 AA.  
 AC W13010;  
 DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.  
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;  
 KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.  
 OS Homo sapiens.  
 PN DE19531033-A1.  
 PD 27-FEB-1997.  
 PE 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.  
 PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;  
 DR WPI: 97-146518/14.  
 PT Antibody reactive with part of desmosomal cadherin - exposed on  
 PT surface of epithelial or carcinoma cells, not bound to desmosomes,  
 PS useful for diagnosis and treatment of carcinoma micrometastases  
 PS Claim 9; Page 5; 8pp; German.  
 CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or  
 CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to  
 CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separate, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to  
 CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or  
 CC carcinomas.  
 SO Sequence 263 AA;

Query Match 56.6%; Score 43; DB 1; Length 263;  
 Best Local Similarity 50.0%; Pred. No. 2.2;  
 Matches 7; Conservative 5; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 PMLLSRNTGEVRT 15  
 Db 58 PVEFLNKDTGEIYT 71

RESULT 10  
 W38513  
 ID W38513 standard; protein; 249 AA.  
 AC W38513;  
 DT 06-NOV-1998 (first entry)  
 DE S. pneumoniae possible cation transporting ATPase PAFL.  
 KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;  
 KW immunological response; inoculation; antibody production; inhibitor;  
 KW T cell immune response; antimicrobial compound; bacterial adhesion;  
 KW extracellular matrix protein; protein-mediated cell invasion; wound;  
 KW pathogenesis.

OS Streptococcus pneumoniae.  
 PN W09743303-A1.  
 PD 20-NOV-1997.  
 PE 14-MAY-1997; U07950.  
 PR 14-MAY-1996; US-017670.  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PI (SMIK ) SMITHKLINE BEECHAM PLC.  
 PI Black MT, Hodgson JE, Knowles DJC, Nicholas RO,  
 DR WPI: 98-008793/01.  
 DR N-PSDB: T98578.  
 PT Novel Streptococcus pneumoniae proteins and related DNA - useful for  
 PT diagnosing anti-microbial agents for treatment of bacterial  
 PT infections  
 PS Claim 12; Page 296; 483pp; English.  
 CC This represents a Streptococcus pneumoniae protein that is a possible  
 CC cation transporting ATPase, and is encoded by a DNA of the invention.  
 CC The DNA sequences were isolated from Streptococcus pneumoniae strain  
 CC 0100993 (NCIMB 40794). The Streptococcus pneumoniae proteins of the  
 CC invention can be used to identify compounds which interact with and  
 CC inhibit or activate the activity of the proteins. Antagonists can be  
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic  
 CC immunisation. They can also be used to induce an immunological response  
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery  
 CC of the encoding nucleic acids in a vector adequate to produce antibody  
 CC and/or T cell immune responses to protect the animal from disease. The  
 CC proteins can also be used to identify antimicrobial compounds which are  
 CC capable of inhibiting their bioactivity. In particular the proteins of  
 CC the invention can be used to prevent adhesion of bacteria to mammalian  
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to  
 CC block protein-mediated mammalian cell invasion, and to block the normal  
 CC progression of pathogenesis in infections initiated other than by the  
 CC implantation of in-dwelling devices or other surgical techniques.  
 SO Sequence 249 AA;

Query Match 51.3%; Score 39; DB 1; Length 249;  
 Best Local Similarity 60.0%; Pred. No. 11;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 TPELLSRNTGEVRT 15  
 Db 201 TIOVLLSANRAEVIYT 215

RESULT 11  
 W80699  
 ID W80699 standard; protein; 835 AA.  
 AC W80699;  
 DT 24-DEC-1998 (first entry)  
 DE S. pneumoniae cation transporting ATPase.  
 KW Streptococcus pneumoniae protein; recombinant; gene expression; DNA chip;  
 KW virulence; antibody; infection; detection; treatment; hypochlorite;  
 KW cell wall biosynthetic, external target; minimal gene set protein.  
 OS Streptococcus pneumoniae.  
 PN W09826072-A1.  
 PD 18-JUN-1998.  
 PE 09-DEC-1997; U22578.  
 PR 13-DEC-1996; US-036281.  
 PA (ELIL ) LILLY & CO ELI.  
 PI Baltz RH, Burgett SG, Dehoff BS, Hoskins JA, Jackunas SR,  
 PI Mills BJ, Norris FH, Peery RB, Rockey PK, Kosteck PR,  
 PI Skatrud PL, Smith MC, Solender PJ, Treadway PJ,  
 DR WPI: 98-348529/30.  
 DR N-PSDB: V65261.  
 PT Streptococcus pneumoniae nucleic acid sequences - used in DNA chips  
 PT for evaluating gene expression, and identification of virulence  
 PT genes  
 PS Claim 3; Pages 280-291; 333pp; English.  
 CC This sequence represents a Streptococcus pneumoniae cation transporting  
 CC ATPase. The invention provides DNA sequences (V65201 to V65304) from  
 CC the Streptococcus pneumoniae genome and corresponding protein sequences

CC (W80605 to W80728). The protein sequences are classified as hypothetical,  
 CC cell wall biosynthetic, external target, or minimal gene set proteins. A  
 CC recombinant host containing a vector comprising any of the above nucleic  
 CC acids can be used for the recombinant expression of the proteins. The  
 CC invention also provides a DNA chip having arrayed on it at least 15 base  
 CC pair fragment of any one or more of these DNA sequences. The DNA chip can  
 CC be used methods for evaluating gene expression in *S. pneumoniae* and for  
 CC identifying virulence genes in *S. pneumoniae*. Antibodies that selectively  
 CC bind to the above proteins or peptide fragments can be used to treat  
 CC *S. pneumoniae* infection. The antibodies can also be used to detect  
 CC *S. pneumoniae* cells.  
 CC Sequence 835 AA;  
 SO

Query Match 51.3%; Score 39; DB 1; Length 835;  
 Best Local Similarity 60.0%; Pred. NO. 41;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 TPMFLSRNTGEVPT 15  
 1 : ||| || ||  
 DB 622 TIOYLSANTAEVLT 636

RESULT 12  
 ID R07998 standard; protein; 586 AA.  
 AC R07998; 26-FEB-1991 (first entry)  
 DE Asparagine synthetase AS1.  
 KW asparagine synthetase; transgenic plant; herbicide resistance;  
 KM drought tolerance; nitrogen fixation; pea.  
 OS Pisum sativum.  
 FH Key Location/Qualifiers  
 FT region 116..128  
 FT region /label= 80% homology with human AS1 sequence  
 FT region 218..243  
 FT region /label= 80% homology with human AS1  
 FT region 340..348  
 FT region /label= 80% homology with human AS1  
 FT region 352..360  
 FT region /label= 80% homology with human AS1  
 FT region 392..401  
 FT region /label= 80% homology with human AS1  
 FT region 486..500  
 FT region /label= 80% homology with human AS1  
 FT binding\_site 1..4  
 FT /label= putative glutamine binding site  
 PN W09013633-A.  
 PD 15-NOV-1990.  
 PE 02-MAY-1990; U02443.  
 PR 03-MAY-1989; US-347302.  
 PR 26-APR-1990; US-514816.  
 PA (UYRO-) ROCKFELLER UNIV.  
 PI Coruzzi GM, Tsai FY.  
 DR WPI: 90-361471/48.  
 DR N-PSDB: 006598.  
 PT Plant asparagine synthetase - includes DNA expression system and  
 PT transgenic plants  
 PS Disclosure; Fig 2A; 91pp; English.  
 CC The DNA sequence encoding this protein was isolated from cDNA  
 CC clones selected from a pea nucleotide cDNA library from the "Sparkle"  
 CC variety of *P. sativum*. Human AS cDNA was used as a probe.  
 CC The protein is produced by expression vectors containing the AS1  
 CC coding sequence. Recombinant AS can be used to engineer herbicide  
 CC resistance, as a dominant selectable marker, to select novel  
 CC herbicides or compounds useful for synchronizing plant cells in  
 CC culture, etc.  
 CC See also 006599, 006622 and 006623.  
 CC Sequence 586 AA;  
 SO

Query Match 50.0%; Score 38; DB 1; Length 586;  
 Best Local Similarity 87.5%; Pred. NO. 43;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 TPMFLSR 8  
 1 : |||||  
 DB 323 TPMFLMSR 330

RESULT 13  
 ID R07999 standard; protein; 583 AA.  
 AC R07999; 26-FEB-1991 (first entry)  
 DE Asparagine synthetase AS2.  
 KW asparagine synthetase; transgenic plant; herbicide resistance;  
 KM drought tolerance; nitrogen fixation; pea.  
 OS Pisum sativum.  
 FH Key Location/Qualifiers  
 FT binding\_site 1..4  
 FT /label= glutamine binding site  
 FT /note= "conserved in human AS and pea AS1 and 2."  
 PN W09013633-A.  
 PD 15-NOV-1990.  
 PE 02-MAY-1990; U02443.  
 PR 03-MAY-1989; US-347302.  
 PR 26-APR-1990; US-514816.  
 PA (UYRO-) ROCKFELLER UNIV.  
 PI Coruzzi GM, Tsai FY.  
 DR WPI: 90-361471/48.  
 DR N-PSDB: 006599.  
 PT Plant asparagine synthetase - includes DNA expression system and  
 PT transgenic plants  
 PS Disclosure; Fig 2B; 91pp; English.  
 CC The DNA sequence encoding this protein was isolated from cDNA  
 CC clones selected from a pea nucleotide cDNA library from the "Sparkle"  
 CC variety of *P. sativum*. Pea AS1 cDNA was used as a probe.  
 CC The protein is produced by expression vectors containing the AS2  
 CC coding sequence. Recombinant AS can be used to engineer herbicide  
 CC resistance, as a dominant selectable marker, to select novel  
 CC herbicides or compounds useful for synchronizing plant cells in  
 CC culture, etc.  
 CC A comparison of pea AS and human AS polypeptides reveals an overall  
 CC homology of 47% at the amino acid level, c.f. 86% between AS1 and  
 CC AS2. There are several regions of high local homology (greater than  
 CC 80%) shared between the pea AS and human AS polypeptides.  
 CC See also 006598, 006622 and 006623.  
 CC Sequence 583 AA;  
 SO

Query Match 50.0%; Score 38; DB 1; Length 583;  
 Best Local Similarity 87.5%; Pred. NO. 42;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 1 : |||||  
 DB 322 TPMFLMSR 329

RESULT 14  
 ID P94879 standard; protein; 407 AA.  
 AC P94879; 10-MAR-1993 (revised)  
 DT 02-JUL-1990 (first entry)  
 DE ORF incorporated within the HIS4 gene.  
 KW yeast promoter; yeast pheromones; FUS-1; BIK-1; HIS-4.  
 OS Saccharomyces cerevisiae.  
 PN W08810308-A.  
 PD 29-DEC-1988.  
 PE 23-JUN-1988; 02129.  
 PR 24-JUN-1987; US-066078.  
 PA (WHIT-) Whitehead Inst.  
 PI Fink GR, Trueheart J, Ellison EA.  
 DR WPI: 89-023850/03.

DR N-PSDB; N93100.  
 PT New DNA fragment contg. protein encoding gene and yeast promoter  
 PT controlled by mating pheromone allowing efficient and  
 PT regulatable expression.  
 PS Claim 5; Fig 4; 51pp; English.  
 CC ORF lies between the FUS 1 or BIK 1 promoter, BIK 1 running 3' to 5'.  
 CC At least one polypeptide may be promoted within a high-copy vector  
 CC induced by a-factor for alpha cells, alpha-factor for a-cells using this  
 CC promoter system.  
 SO Sequence 407 AA;

Query Match 48.7%; Score 37; DB 1; Length 407;  
 Best Local Similarity 57.1%; Pred. No. 44;  
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 2 PMFLSRNTGEVRT 15  
 DB 366 PMSTEDRNTAETRT 379

## RESULT 15

W62750  
 ID W62750 standard; Protein; 218 AA.  
 AC W62750;  
 DT 09-NOV-1998 (first entry)  
 DE Streptococcus pneumoniae polypeptide.  
 KW Polypeptide; ORF; open reading frame; infection; bacterial;  
 KM streptococcal; bacteremia; diagnosis; prophylaxis.  
 OS Streptococcus pneumoniae.  
 PN W09823631-A1.  
 PD 04-JUN-1998.  
 PF 24-NOV-1997; U21976.  
 PR 27-NOV-1996; US-031879.  
 PA (SMIX ) SMITHKLINE BEECHAM CORP.  
 PI (SMIX ) SMITHKLINE BEECHAM PLC.  
 PI Black Mt, Hodgson JE, Knowles DGC, Lonetto MA, Nicholas RO,  
 PI Reid RH, Zarfos PN;  
 DR WPI: 98-322654/28.  
 PT Streptococcus pneumoniae polynucleotides - useful for developing  
 PT products for diagnosis, prevention and treatment of infections e.g.  
 PT pneumonia, bacteremia, meningitis or endocarditis  
 PS Claim 5; Page 32; 181pp; English.  
 CC The sequence is that of a Streptococcal polypeptide.  
 CC The polypeptide can potentially be used for the diagnosis and  
 CC prevention of bacterial infections, especially SP infection.  
 CC It may be used for the treatment of diseases such as otitis media,  
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural  
 CC empyema, endocarditis or infection of the cerebrospinal fluid.  
 SO Sequence 218 AA;

Query Match 48.7%; Score 37; DB 1; Length 218;  
 Best Local Similarity 38.9%; Pred. No. 22;  
 Matches 7; Conservative 6; Mismatches 1; Indels 4; Gaps 1;

OY 2 PMF-----LLSRNTGEVRT 15  
 DB 44 PIVTFRRLNKRTGEIKT 61

Search completed: January 6, 2000, 15:40:37  
 Job time: 1684 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:21 ; Search time 59.21 Seconds  
(without alignments)  
15.518 Million cell updates/sec

Title: US-08-991-628-4  
Sequence: 1 TPMFLSRMTGEVRT 15

Scoring table:

BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL\_11.\*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	81.6	993	11	035902
2	48	63.2	820	4	015039
3	47	61.8	336	5	009661
4	43.5	57.2	291	2	004824
5	40	52.6	452	1	029889
6	39	51.3	393	2	008395
7	39	51.3	163	2	084262
8	39	51.3	215	5	044017
9	38	50.0	810	4	075279
10	38	50.0	1115	4	095846
11	38	50.0	585	10	065329
12	38	50.0	586	10	040328
13	38	50.0	586	10	P93618
14	38	50.0	581	10	P93167
15	38	50.0	586	10	024483
16	38	50.0	948	10	081211
17	38	50.0	578	10	092877
18	38	50.0	578	10	092876
19	38	50.0	584	10	092823
20	37	48.7	231	2	P71027
21	37	48.7	211	2	083378
22	37	48.7	426	2	092866
23	37	48.7	557	3	042902
24	37	48.7	398	3	060145
25	37	48.7	407	3	094079

26	37	48.7	816	4	075284	075284 homo sapien
27	37	48.7	224	10	004149	004149 arabidopsis
28	37	48.7	367	10	092870	092870 arabidopsis
29	37	48.7	593	12	010288	010288 orgyia pseu
30	36	47.4	466	2	P77286	P77286 escherichia
31	36	47.4	922	4	060330	060330 homo sapien
32	36	47.4	807	2	075277	075277 homo sapien
33	36	47.4	814	4	075281	075281 homo sapien
34	36	47.4	803	4	075283	075283 homo sapien
35	36	47.4	878	4	015855	015855 homo sapien
36	36	47.4	414	5	021405	021405 caenorhabdi
37	36	47.4	643	5	018290	018290 caenorhabdi
38	36	47.4	1417	10	048697	048697 arabidopsis
39	36	47.4	483	10	049927	049927 pisum sativ
40	36	47.4	928	11	008963	008963 rattus norv
41	36	47.4	308	11	035100	035100 rattus norv
42	36	47.4	469	11	062735	062735 rattus norv
43	36	47.4	379	11	060643	060643 mus musculu
44	36	47.4	484	11	060641	060641 mus musculu
45	36	47.4	914	11	088192	088192 mus musculu

## ALIGNMENTS

RESULT 1  
ID 035902 PRELIMINARY; PRT; 993 AA.  
AC 035902;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE DESMOLEIN 3 (FRAGMENT).  
GN D5G3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Euthera; Rodentia; Sciurognath; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RA ISHIKAWA H., LI K., UETTO J.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL: U86016; AAB65091.1; -;  
DR PFM: PF00028; cadherin; 4.  
DR PROSITE: PS00232; CADHERIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON\_TER 993 993  
FT SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 81.6%; Score 62; DB 11; Length 993;  
Best Local Similarity 92.3%; Pred. No. 0.0024;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 FELLSRMTGEVRT 15  
Db 208 MFLSRMTGEVRT 220  
OY  
Db  
RESULT 2  
ID 015039 PRELIMINARY; PRT; 820 AA.  
AC 015039;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE KIAA0327 PROTEIN.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Euthera; Primates; Catarrhini; Homnidae; Homo.  
RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 97349984.  
 RA NAGASE T., ISHIKAWA K., NAKAJIMA D., CHIRA M., SEKI N., MIYAJIMA N., TANAKA A., KOTANI H., NOKURA N., OHARA O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro."  
 RL DNA Res. 4:141-150(1997).  
 CC EMBL: AB002325; BAA20785.1; -  
 DR PFAM: PF00028; cadherin; 5.  
 DR PROSITE: PS00232; CADHERIN; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 820 AA; 89864 MW; 386CA92 CRC32;

Query Match 63.2%; Score 48; DB 4; Length 820;  
 Best Local Similarity 61.5%; Pred. No. 0.84;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPEFLSRNTGEV 13  
 |||:|:||||  
 Db 289 TPLFLNNTGEI 301

RESULT 3  
 ID 009661 PRELIMINARY; PRT; 336 AA.  
 AC 009661;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE HYPOTHEICAL 37.7 KD PROTEIN ZK177.6 IN CHROMOSOME II.  
 GN ZK177.6.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;  
 OC Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA ANDERSON K.;  
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.  
 CC 1- SIMILARITY: WEAK, TO YEAST PROTEIN CDC20.  
 DR EMBL: U21321; AAB36970.1; -  
 DR WORMPEP: ZK177.6; CE02095.  
 DR PFAM: PF00400; WD40; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 336 AA; 37668 MW; 2DC77B26 CRC32;

Query Match 61.8%; Score 47; DB 5; Length 336;  
 Best Local Similarity 57.1%; Pred. No. 0.49;  
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 PMLFSRNTGEV 15  
 |||:|:||||  
 Db 187 PMLFIDHTGQV 200

RESULT 4  
 ID 004824 PRELIMINARY; PRT; 291 AA.  
 AC 004824;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
 DE ORF6 PROTEIN PRECURSOR.  
 OS Salmonella typhimurium.  
 OS Plasmid 90 kb virulence.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2;  
 RX MEDLINE: 93316852.  
 RA FRIEDRICH M.J., KINSEY N.E., VILA J., KADNER R.J.;  
 RT "Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence plasmid of *Salmonella typhimurium*: the presence of fimbrial biosynthetic genes."  
 RL Mol. Microbiol. 8:543-558(1993).  
 DR EMBL: L08613; AAC36963.1; -  
 KW Plasmid; Signal.  
 FT SIGNAL  
 FT CHAIN 1  
 SQ SEQUENCE 291 AA; 30683 MW; CAD5ADDF CRC32;

Query Match 57.2%; Score 43.5; DB 2; Length 291;  
 Best Local Similarity 80.0%; Pred. No. 1.9;  
 Matches 12; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

OY 1 TPEFLSRNTGEV 15  
 |||:|:||||  
 Db 128 TLMFSLR-TGEV 141

RESULT 5  
 ID 029889 PRELIMINARY; PRT; 452 AA.  
 AC 029889;  
 DT 01-JAN-1998 (TREMBlrel. 05, Created)  
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE DNA REPAIR PROTEIN RAD25.  
 GN AF0358.  
 OS Archaeoglobus fulgidus.  
 OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;  
 OC Archaeoglobus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE: 96049343.  
 RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E., KETCHUM K.A., DODSON R.J., GINN M., HICKEY E.K., PETERSON J.D., RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KRAPIDES N.C., FLEISCHMAN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S., KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTIS B., PETERSON S., REICH C.I., MCNEIL L.K., BAUGER J.H., GLODER A., ZHOU L., OVERBEER R., GOCAYNE J.D., WEIDMAN J.F., McDONALD L., UTTERBACK T., RA COTTON M.D., SPRIGGS T., ARTIACH P., KAINE B.P., SYKES S.M., SADOW P.W., D'ANDREA K.F., BOWMAN C., FUJII C., GARLAND S.A., MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., NOSE C.R., VENTER J.C.;  
 RA "The complete genome sequence of the hyperthermophilic, RT sulfate-reducing archaeon *Archaeoglobus fulgidus*."  
 RL Nature 390:364-370(1997).  
 DR EMBL: AEO01080; AAB90879.1; -  
 DR TIGR: AF0358; -  
 DR PFAM: PF00270; DEAD; 1.  
 DR PFAM: PF00271; helicase-C; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 452 AA; 51768 MW; 615C80A2 CRC32;

Query Match 52.6%; Score 40; DB 1; Length 452;  
 Best Local Similarity 72.7%; Pred. No. 14;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 LISRNTGEV 15  
 |||:|:||||  
 Db 430 LISRGTEVNT 440

RESULT 6  
 ID 008395

ID 008395 PRELIMINARY: PRT: 393 AA.  
AC 008395:  
DT 01-JUL-1997 (TREMBLrel. 04, Created)  
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE CITRATE SYNTHASE (EC 4.1.3.7).  
GN GLT3.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN-H37RV;  
RA DEVIN K., CHURCHER C.M.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RA PARKHILL J., BARRELL B.G., RAINDREAN M.A.;  
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE: 96181548.  
RA PHILIPP W.J., POULET S., EIGMEIER K., PASCOPELLA L.,  
RA BALASUBRAMANIAN V., HEYM B., BERGH S., BLOOM B.R., JACOBS W.R. JR.,  
RA COLE S.T.;  
RT "An integrated map of the genome of the tubercle bacillus,  
RT Mycobacterium tuberculosis H37RV, and comparison with Mycobacterium  
RT leprae.";  
RT Proc. Natl. Acad. Sci. U.S.A. 93:3132-3137(1996).  
CC -1- CITRATE SYNTHASE IS FOUND IN NEARLY ALL CELLS CAPABLE OF  
CC -1- OXIDATIVE METABOLISM.  
CC -1- CATALYTIC ACTIVITY: CITRATE + COA - ACETYL-COA + H(2)O +  
CC OXALOACETATE.  
CC -1- PATHWAY: TRICARBOXYLIC ACID CYCLE.  
CC -1- SIMILARITY: TO OTHER CITRATE SYNTHASES.  
DR EMBL: 295585; CAB09042.1; -  
DR PROSITE: P800480; CITRATE\_SYNTHASE. 1.  
DR PFM: PF00285; citrate\_synth. 1.  
KW Lyase; tricarboxylic acid cycle.  
SQ SEQUENCE 393 AA; 42969 MW; 8850A8D3 CRC32;

Query Match 51.3%; Score 39; DB 2; Length 393;  
Best Local Similarity 63.8%; Pred. No. 19;  
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 TPELLSRNTG 11  
||:|||||  
DB 350 TPIFWMSRTG 360

RESULT 7  
ID 084262 PRELIMINARY: PRT: 163 AA.  
AC 084262:  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE HYPOTHETICAL 18.9 KD PROTEIN.  
GN CT260.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-D/WR-3/CX;  
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,  
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,  
RA DAVIS R.W.;  
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:  
RT Chlamydia trachomatis.";  
RL Science 0:0-0(1998).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-D/WR-3/CX;  
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,  
RA MITCHELL W.P., OLINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,  
RA DAVIS R.W.;  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AE001299; AAC67853.1;  
KM Hypothetical protein.  
SQ SEQUENCE 163 AA; 18875 MW; 80D724D1 CRC32;

Query Match 51.3%; Score 39; DB 2; Length 163;  
Best Local Similarity 58.3%; Pred. No. 7.3;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 PPELLSRNTGEV 13  
||:|||||  
DB 43 PLFVIRNNGEV 54

RESULT 8  
ID 044017 PRELIMINARY: PRT: 215 AA.  
AC 044017:  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE ERD2.  
GN ERD2.  
OS Entamoeba histolytica.  
OC Eukaryota; Entamoebidae; Entamoeba.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-HK-9;  
RA SANCHEZ-LOPEZ R., GAMA-CASTRO S., RAMOS M.A., MERINO E., LIZARDI P.M.,  
RA ALAGON A.;  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ002138; CAA05206.1;  
DR PFM: PF00810; ER\_lumen\_recept. 1.  
KW SEQUENCE 215 AA; 25724 MW; 28FF1C82 CRC32;

Query Match 51.3%; Score 39; DB 5; Length 215;  
Best Local Similarity 66.7%; Pred. No. 9.8;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PPELLSRNTGEV 13  
||:|||||  
DB 135 POLLLSRSTGEV 146

RESULT 9  
ID 075279 PRELIMINARY: PRT: 810 AA.  
AC 075279:  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE KIA0345-LIKE 3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA KIMBERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,  
RA KADNER K., MIGUEL T., MILLER C., PITLUCK S., POLLARD M., ROJESKI R.,  
RA SUBRAMANIAN S., MARIN C.H.;  
RT "Sequencing of human chromosome 5";  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA RICKE D.O.;



RT "Large Scale Sequence Analysis and Annotation with the Sequence  
 Comparison Analysis (SCAN) System."  
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AC005609; AAC34316.1; -  
 DR PFAM: PF00028; cadherin; 6  
 DR PROSITE: PS00232; CADHERIN; 5  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 810 AA; 88391 MW; 9FAE3C7D CRC32;

Query Match 50.0%; Score 38; DB 4; Length 810;  
 Best Local Similarity 58.3%; Pred. No. 64;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 3 TPFLLSRNTGEVR 14  
 DB 291 LFTLDQNGEVR 302

RESULT 10  
 OY5846 PRELIMINARY; PRT; 1115 AA.  
 ID OY5846 -  
 AC OY5846;  
 DT 01-MAY-1999 (TREMblrel. 10, Created)  
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE EUKARYOTIC TRANSLATION INITIATION FACTOR 2 ALPHA KINASE PEK.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Euthera; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PANCREAS;  
 RX MEDLINE; 99150360.  
 RA SHI Y., AN J., LIANG J., HAYES S.E., SANDUSKY G.E., STRAM L.E.,  
 RT "Characterization of a Mutant Pancreatic eIF-2alpha Kinase, PEK, and  
 RT Co-localization with Somatostatin in Islet Delta Cells."  
 RL J. Biol. Chem. 274:5723-5730(1999).  
 DR EMBL: AF101046; AAD1961.1;  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Kinase; Initiation factor.  
 SQ SEQUENCE 1115 AA; 125146 MW; 5BB6FC8 CRC32;

Query Match 50.0%; Score 38; DB 4; Length 1115;  
 Best Local Similarity 53.3%; Pred. No. 90;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 TPFLLSRNTGEVRT 15  
 DB 972 TPFYARHTGOVGT 986

RESULT 11  
 OY5329 PRELIMINARY; PRT; 585 AA.  
 ID OY5329 -  
 AC OY5329;  
 DT 01-AUG-1998 (TREMblrel. 07, Created)  
 DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE ASPARAGINE SYNTHETASE.  
 GN AS.  
 OS Eleagnus umbellata.  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC eumphylophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Rosales; Elaeagnaceae; Elaeagnus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-ROOT NODULE;  
 RA KIM H.-B., AN C.-S.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF061740; AAC16325.1; -  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PFAM: PF00310; GATase\_2; 1.  
 SQ SEQUENCE 585 AA; 65878 MW; 8A12777D CRC32;

Query Match 50.0%; Score 38; DB 10; Length 585;  
 Best Local Similarity 87.5%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFLLSR 8  
 DB 322 TPFLLSR 329

RESULT 12  
 OY40328 PRELIMINARY; PRT; 586 AA.  
 ID OY40328 -  
 AC OY40328;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE ASPARAGINE SYNTHETASE.  
 OS Medicago sativa (Alfalfa).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC eumphylophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 CC Medicago.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. SARANAC;  
 RA GANTT S.;  
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: L40327; AAB48058.1; -  
 DR MENDEL; 9036; MEDSA; 1042; 1.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PFAM: PF00310; GATase\_2; 1.  
 SQ SEQUENCE 586 AA; 66462 MW; 962C3A3F CRC32;

Query Match 50.0%; Score 38; DB 10; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPFLLSR 8  
 DB 323 TPFLLSR 330

RESULT 13  
 OY53618 PRELIMINARY; PRT; 586 AA.  
 ID OY53618 -  
 AC OY53618;  
 DT 01-MAY-1997 (TREMblrel. 03, Created)  
 DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE ASPARAGINE SYNTHETASE (EC 6.3.5.4) (ASPARGINE SYNTHASE  
 DE (GLUTAMINE-HYDROLYSING)).  
 GN VPAS1.  
 OS Vicia faba (Broad bean).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC eumphylophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 CC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 CC Vicia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-KLEINE THUNDER; TISSUE-ROOT NODULE;  
 RA KOESTER H., ALBOS O., FRIEDLING M., TCHETKOVA S.A., TIKHONOVITCH I.A.;  
 RA PUEHLER A., PERLICK A.M.;  
 RL Plant Sci. 124:89-95(1997).  
 CC -1- CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +  
 CC DIPHOSPHATE + L-ASPARGINE + L-GLUTAMATE.

DR EMBL: 272354; CAA96526.1; -  
 DR MENDEL: 11023; VTCfa;1042.1.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PFAM: PF00310; GATase\_2; 1.  
 KW L1gase.  
 SQ SEQUENCE 586 AA; 66236 MW; AD8BD927 CRC32;

Query Match 50.0%; Score 38; DB 10; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 |||||:  
 Db 323 TPMFLMSR 330

RESULT 14  
 P93167 PRELIMINARY; PRT; 581 AA.

AC P93167;  
 DT 01-MAY-1997 (TREMblrel. 03; Created)  
 DT 01-MAY-1997 (TREMblrel. 03; Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10; Last annotation update)  
 DE ASPARAGINE SYNTHETASE 2 (EC 6.3.5.4) (ASPARGINE SYNTHASE  
 DE (GLUTAMINE-HYDROLYSING)) (ASPARGINE SYNTHETASE  
 GN AS2.  
 OS Glycine max (Soybean).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
 OC euphyllophytes: Spermatophyta: Magnoliophyta; eudicotyledons;  
 OC core eudicots; rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 OC Glycine.  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CENTURY; TISSUE-MATURE LEAF;  
 RX MEDLINE: 97188563.  
 RA HUGHES C.A., BEARD H., MATTHEWS B.F.;  
 RT "Molecular cloning and expression of two cDNAs encoding asparagine  
 RT synthetase in soybean."  
 RL Plant Mol. Biol. 33:301-311(1997).  
 CC -1 CATALYTIC ACTIVITY: ATP + L-ASPARTATE + L-GLUTAMINE -> AMP +  
 CC DIPHOSPHATE + L-ASPARGINE + L-GLUTAMATE.  
 DR EMBL: U77678; AAC49613.1; -  
 DR MENDEL: 81441; GLYma;1042.1.  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PFAM: PF00310; GATase\_2; 1.  
 KW L1gase.  
 SQ SEQUENCE 581 AA; 65609 MW; 90AF9E2F CRC32;

Query Match 50.0%; Score 38; DB 10; Length 581;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 |||||:  
 Db 322 TPMFLMSR 329

RESULT 15  
 024483

ID 024483 PRELIMINARY; PRT; 586 AA.

AC 024483;  
 DT 01-JAN-1998 (TREMblrel. 05; Created)  
 DT 01-JAN-1998 (TREMblrel. 05; Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10; Last annotation update)  
 DE ASPARAGINE SYNTHETASE  
 OS Medicago sativa (Alfalfa).  
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta; Tracheophyta;  
 OC euphyllophytes: Spermatophyta: Magnoliophyta; eudicotyledons;  
 OC core eudicots; rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;  
 OC Medicago.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97432147.  
 RA SHI L., TWARDY S.N., YOSHIOKA H., GREGGSON R.G., MILLER S.S.,  
 RA SAMAC D.A., GANTT J.S., UNKEFER P.J., VANCE C.P.;  
 RT "Nitrogen assimilation in alfalfa: Isolation and characterization of  
 RT an asparagine synthetase gene showing enhanced expression in root  
 RT nodules and dark-adapted leaves."  
 RL Plant Cell 9:1339-1356(1997).  
 DR EMBL: U89923; AAB81011.1; -  
 DR PFAM: PF00733; Asn\_synthase; 1.  
 DR PFAM: PF00310; GATase\_2; 1.  
 SQ SEQUENCE 586 AA; 66448 MW; 63C8B6BA CRC32;

Query Match 50.0%; Score 38; DB 10; Length 586;  
 Best Local Similarity 87.5%; Pred. No. 45;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 TPMFLSR 8  
 |||||:  
 Db 323 TPMFLMSR 330

Search completed: January 12, 2000, 23:15:22  
 Job time: 185 sec

Sat Jan 15 11:45:02 2000

us-08-991-628-4.rpt

Page 6

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:42 ; Search time 65.16 Seconds  
(without alignments)  
10,858 Million cell updates/sec

Title: US-08-991-628-5  
Perfect score: 88  
Sequence: 1 CECNIKVDVNDNFP 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR\_62:\*

Word size: 0

Number of hits that pass the threshold : 142080

1: PIR1:\*\n2: PIR2:\*\n3: PIR3:\*\n4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	100.0	999	1	desmoglein 3 precu
2	74	84.1	1049	1	desmoglein 1 precu
3	74	84.1	1043	1	desmoglein 1 precu
4	52	59.1	832	2	LI-cadherin - huma
5	51	58.0	2163	2	hypothetical prote
6	49	55.7	863	1	desmocollin 2a pre
7	49	55.7	809	1	desmocollin 2b pre
8	48	54.5	809	1	cadherin-related t
9	48	54.5	809	1	cadherin-5 precurs
10	47	53.4	896	2	desmocollin - bov1
11	47	53.4	3343	2	ZK112.7 protein -
12	46	52.3	1117	2	desmoglein 2 - hum
13	46	52.3	793	2	cadherin 8 - huma
14	46	52.3	793	2	cadherin-14 - huma
15	46	52.3	790	2	desmocollin 68 -
16	45	51.1	901	1	desmocollin 3a pre
17	45	51.1	847	1	desmocollin 3b pre
18	45	51.1	794	2	cadherin 12 - huma
19	45	51.1	805	2	translatin initia
20	45	51.1	656	2	serine/threonine-s
21	45	51.1	827	2	LI-cadherin precu
22	44	50.0	884	1	E-cadherin precurs
23	44	50.0	884	1	uvomorulin - mouse
24	44	50.0	790	2	cadherin-6 - huma
25	44	50.0	789	2	K-cadherin - rat
26	44	50.0	785	2	cadherin-7 - chick
27	44	50.0	790	2	cadherin-6s - chic
28	44	50.0	790	2	F-cadherin - Afri
29	44	50.0	807	2	potassium channel
30	44	50.0	787	2	MEG1 protein - ra
31	44	50.0	4351	2	proliferating-cell
32	44	50.0	682	2	hypothetical prote
33	44	50.0	620	2	hypothetical prote
34	44	50.0	3051	2	DN-cadherin - frul
35	44	50.0	3097	2	

36 44 50.0 829 2 146536  
37 44 50.0 235 2 E70378  
38 43 48.9 796 2 A38982  
39 43 48.9 796 2 148277  
40 43 48.9 796 2 149556  
41 43 48.9 796 2 A53584  
42 42 48.9 601 2 A27020  
43 42.5 48.3 135 2 S63377  
44 42 47.7 285 2 B71194  
45 41 46.6 394 2 H64448

## ALIGNMENTS

RESULT 1  
desmoglein 3 precursor - human  
N:Alternate names: pemphigus vulgaris antigen  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: A41088  
R:Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.  
Cell 67, 869-877, 1991  
A>Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a  
A:Reference number: A41088; MUID:92069753  
A:Accession: A41088  
A:Molecule type: mRNA  
A:Residues: 1-999 <MAN>  
A:Cross-references: GB:M76482; NID:9190751; PIDN:AAA60230.1; PID:9190752  
C:Gene: GDB:DSG3  
A:Cross-references: GDB:134030; OMIM:169615  
A:Map position: 16q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-999/Product: desmoglein homolog #status predicted <MAT>  
F:50-615/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR2>  
F:160-267/Domain: cadherin repeat homology <CR3>  
F:270-383/Domain: cadherin repeat homology <CR4>  
F:390-495/Domain: cadherin repeat homology <CR5>  
F:496-598/Domain: cadherin repeat homology <CR6>  
F:616-635/Domain: transmembrane #status predicted <TM>  
F:640-999/Domain: intracellular #status predicted <INT>  
F:910-938/Domain: desmoglein repeat <DG1>  
F:937-966/Domain: desmoglein repeat <DG2>  
F:110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 88; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 2.1e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CECNIKVDVNDNFP 15  
DB 251 CECNIKVDVNDNFP 265

RESULT 2  
desmoglein 1 precursor - human  
N:Alternate names: desmosomal glycoprotein 1  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: S16906; A39706; A61254; A61279; S16158  
R:Buxton, R.S.  
submitted to the EMBL-Data Library, November 1990  
A:Reference number: S16906  
A:Accession: S16906  
A:Molecule type: mRNA

A:Residues: 1-1049 <BX>  
A:Cross-references: EMBL:X56654; NID:930505; PIDN:CAA39976.1; PID:930506  
R:Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arneemann, J.; R  
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991  
A:Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,  
A:Reference number: A39706; MUID:91271279  
A:Accession: A39706  
A:Molecule type: mRNA  
A:Residues: 24-1049 <WHE>  
A:Cross-references: GB:X56654  
R:Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.  
J. Cell Sci. 99, 809-821, 1991  
A:Title: Structural analysis and expression of human desmoglein: a cadherin-like compone  
A:Reference number: A61254; MUID:92121251  
A:Accession: A61254  
A:Molecule type: mRNA  
A:Residues: 26-1049 <NML>  
R:Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; Kling, I.A.; Magee  
Biochem. Soc. Trans. 19, 1060-1064, 1991  
A:Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily  
A:Reference number: A61279; MUID:92175187  
A:Accession: A61279  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-55 <WH3>  
C:Genetics:  
A:Gene: GDB:DSG1  
A:Cross-references: GDB:126563; OMIM:125670  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro  
F:1-73/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-1049/Product: desmoglein #status predicted <MAT>  
F:52-157/Domain: extracellular #status predicted <EXT>  
F:160-269/Domain: cadherin repeat homology <CR1>  
F:272-385/Domain: cadherin repeat homology <CR2>  
F:392-493/Domain: cadherin repeat homology <CR3>  
F:509-530/Region: serine/threonine-rich  
F:549-569/Domain: transmembrane #status predicted <TM>  
F:572-1049/Domain: intracellular #status predicted <INT>  
F:840-869/Domain: desmoglein repeat <DG1>  
F:870-899/Domain: desmoglein repeat <DG2>  
F:900-927/Domain: desmoglein repeat <DG3>  
F:928-956/Domain: desmoglein repeat <DG4>  
F:969-1019/Region: glycine/serine-rich  
F:110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.1%; Score 74; DB 1; Length 1049;  
Best Local Similarity 80.0%; Pred. No. 0.0004;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CECNIKVDNDNFP 15  
|||||: |||||  
DB 253 CECNIKVDNDNFP 267

RESULT 3  
13BOG1  
desmoglein 1 precursor - bovine  
N:Alternate names: desmoglein BDGM  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: S14603; A38872; A37785; S24412  
R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.  
submitted to the EMBL Data Library March 1991  
A:Description: Complete sequence of the desmoglein precursor and evidence for the existe  
A:Reference number: S14603  
A:Accession: S14603  
A:Molecule type: mRNA  
A:Residues: 1-1043 <KOC>

A:Cross-references: EMBL:X58466; NID:9306; PIDN:CAA41380.1; PID:9307  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Eur. J. Cell Biol. 55, 200-208, 1991  
A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypepti  
A:Reference number: A38872; MUID:92037656  
A:Accession: A38872  
A:Molecule type: mRNA  
A:Residues: 1-87; 968-1043 <KOC2>  
A:Cross-references: GB:S64268; GB:S64270  
R:Goddyn, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.  
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990  
A:Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion  
A:Reference number: A37785; MUID:91097553  
A:Accession: A37785  
A:Molecule type: mRNA  
A:Residues: 44-123; 'V', 125-493 <GOO>  
A:Cross-references: GB:M58165; NID:9162966; PIDN:AAA62709.1; PID:9552318  
R:Zimbelmann, R.  
submitted to the EMBL Data Library, February 1991  
A:Reference number: S38721  
A:Accession: S38721  
A:Molecule type: mRNA  
A:Residues: 44-1043 <ZIK>  
A:Cross-references: EMBL:X57784; NID:9436061; PIDN:CAA40930.1; PID:9436062  
R:Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.  
Eur. J. Cell Biol. 55, 1-12, 1990  
A:Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m  
A:Reference number: A48173; MUID:91168965  
A:Accession: A48173  
A:Molecule type: mRNA  
A:Residues: 44-1001; 'AOPPSAT' <KOC3>  
A:Cross-references: GB:X57784  
A:Note: this sequence has been revised in references A38872 and S38721  
C:Genetics:  
A:Gene: DSG1  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-73/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-1043/Product: desmoglein #status predicted <MAT>  
F:52-157/Domain: extracellular #status predicted <EXT>  
F:160-269/Domain: cadherin repeat homology <CR1>  
F:272-385/Domain: cadherin repeat homology <CR2>  
F:392-491/Domain: cadherin repeat homology <CR3>  
F:549-574/Domain: transmembrane #status predicted <TM>  
F:575-1043/Domain: intracellular #status predicted <INT>  
F:846-875/Domain: desmoglein repeat <DG1>  
F:876-905/Domain: desmoglein repeat <DG2>  
F:906-933/Domain: desmoglein repeat <DG3>  
F:934-962/Domain: desmoglein repeat <DG4>  
F:963-1012/Region: glycine/serine-rich  
F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 84.1%; Score 74; DB 1; Length 1043;  
Best Local Similarity 80.0%; Pred. No. 0.0004;  
Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 CECNIKVDNDNFP 15  
|||||: |||||  
DB 253 CECNIKVDNDNFP 267

RESULT 4  
55396  
li-cadherin - human  
C:Species: Homo sapiens (man)  
C:Date: 15-Jul-1995 #sequence\_revision 01-Sep-1995 #text\_change 10-Sep-1997  
R:Boettlinger, A.; Krefl, B.; Fieger, C.; Dlouhy, B.; Berndorff, D.; Goessner, R.; Tau  
submitted to the EMBL Data Library, December 1994

A:Description: Molecular cloning of human LI-cadherin: evidence for a novel type of cadherin  
A:Reference number: S55396  
A:Accession: S55396  
A>Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-832 <BOE>  
A:Cross-references: EMBL:X83228; NID:9854174; PID:9854175  
C:Superfamily: cadherin repeat homology  
F:455-566/Domain: cadherin repeat homology <CR3>

Query Match 59.1%; Score 52; DB 2; Length 832;  
Best Local Similarity 64.3%; Pred. No. 1.1;  
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 ECNIKVKDNDNFP 15  
| : : : : : |  
Db 325 EIHVKVDINDNFP 338

RESULT 5  
T15276  
hypothetical protein R10F2.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T15276  
R:Du, Z.; Gattung, S.  
Submitted to the EMBL Data Library, May 1997  
A:Description: The sequence of C. elegans cosmid R10F2.  
A:Reference number: 218320  
A:Accession: T15276  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2163 <DUZ>  
A:Cross-references: EMBL:AF003388; NID:92088850; PID:92088852; PIDN:AA54266.1; GSPDB:GN  
A:Experimental source: strain Bristol N2; clone R10F2  
C:Genetics:  
A:Gene: CESP.R10F2.1  
A:Map position: 3  
A:Insertions: 34/1; 114/3; 735/3; 1552/2; 1800/3; 1865/3; 1911/3; 2026/3; 2108/2; 2150/3

Query Match 58.0%; Score 51; DB 2; Length 2163;  
Best Local Similarity 53.3%; Pred. No. 4.1;  
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 CECNIKVKDNDNFP 15  
| : : : : : |  
Db 1629 CKCHVYLDNDNLP 1643

RESULT 6  
IJR0DC  
desmocollin 2a precursor - bovine (fragment)  
N:Alternate names: epithelial type 2 desmocollin subform I  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: A41799  
R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky, R.; Franke, W.W.  
Proc. Natl. Acad. Sci. U.S.A. 89, 353-357, 1992  
A:Title: Complexity and expression patterns of the desmosomal cadherins.  
A:Reference number: A41799; MUID:92108053  
A:Accession: A41799  
A:Molecule type: mRNA  
A:Residues: 1-863 <KOC>  
A:Cross-references: GB:M81190; NID:9163757; PIDN:AAA30782.1; PID:9163758  
A:Experimental source: muzzle  
A>Note: sequence extracted from NCBI backbone  
A>Note: 264-Gln and 333-Gln were also found  
C:Genetics:  
A:Gene: DSC2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprotein

F:1-89/Domain: propeptide #status predicted <PRO>  
F:90-663/Product: desmocollin 2a #status predicted <MAT>  
F:90-645/Domain: extracellular #status predicted <EXT>  
F:92-197/Domain: cadherin repeat homology <CR1>  
F:200-309/Domain: cadherin repeat homology <CR2>  
F:312-423/Domain: cadherin repeat homology <CR3>  
F:426-526/Domain: cadherin repeat homology <CR4>  
F:527-634/Domain: cadherin repeat homology <CR5>  
F:646-672/Domain: transmembrane #status predicted <TM>  
F:673-863/Domain: intracellular #status predicted <INT>  
F:120,346,495,579/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:826/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 55.7%; Score 49; DB 1; Length 863;  
Best Local Similarity 61.5%; Pred. No. 3.5;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 CNIKVKDNDNFP 15  
| : : : : : |  
Db 295 CIINIEDVNDNLP 307

RESULT 7  
IJR0DC  
desmocollin 2b precursor - bovine (fragment)  
N:Alternate names: epithelial type 2 desmocollin subform II  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: B41799  
R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Troyanovsky, R.; Franke, W.W.  
Proc. Natl. Acad. Sci. U.S.A. 89, 353-357, 1992  
A:Title: Complexity and expression patterns of the desmosomal cadherins.  
A:Reference number: A41799; MUID:92108053  
A:Accession: B41799  
A:Molecule type: mRNA  
A:Residues: 1-809 <KOC>  
A:Cross-references: GB:M81190; NID:9163757; PIDN:AAA30783.1; PID:9163759  
A:Experimental source: muzzle  
A>Note: sequence extracted from NCBI backbone  
A>Note: 264-Gln and 333-Gln were also found  
C:Genetics:  
A:Gene: DSC2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprotein  
F:1-89/Domain: propeptide #status predicted <PRO>  
F:90-809/Product: desmocollin 2b #status predicted <MAT>  
F:90-645/Domain: extracellular #status predicted <EXT>  
F:92-197/Domain: cadherin repeat homology <CR1>  
F:200-309/Domain: cadherin repeat homology <CR2>  
F:312-423/Domain: cadherin repeat homology <CR3>  
F:426-526/Domain: cadherin repeat homology <CR4>  
F:527-634/Domain: cadherin repeat homology <CR5>  
F:646-672/Domain: transmembrane #status predicted <TM>  
F:673-809/Domain: intracellular #status predicted <INT>  
F:120,346,495,579/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 55.7%; Score 49; DB 1; Length 809;  
Best Local Similarity 61.5%; Pred. No. 3.3;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 CNIKVKDNDNFP 15  
| : : : : : |  
Db 295 CIINIEDVNDNLP 307

RESULT 8  
IJR0DC  
cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Feb-1997  
C:Accession: A41087; B41087

R:Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biesmann, H.; Bryant, P.J.; Goodman, C.S.  
 Cell 67, 853-868, 1991  
 A:Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadherin family  
 A:Reference number: A41087, MUID:92069752  
 A:Accession: A41087  
 A:Molecule type: mRNA  
 A:Residues: 143-485;1279-5147 <MAH>  
 A:Cross-references: GB:M80537  
 A:Accession: B41087  
 A:Molecule type: DNA  
 A:Residues: 1-142;487-1278 <MA2>  
 A:Cross-references: GB:M80537  
 A:Note: 1229-Gly and 1233-Ser were also found  
 C:Genetics:  
 A:Gene: fat  
 A:Cross-references: FlyBase:FBgn0001075  
 C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology  
 C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein  
 F:1-35/Domain: signal sequence #status predicted <SIG>  
 F:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>  
 F:36-4583/Domain: extracellular #status predicted <EXT>  
 F:51-156/Domain: cadherin repeat homology <CR1>  
 F:159-270/Domain: cadherin repeat homology <CR2>  
 F:271-382/Domain: cadherin repeat homology <CR3>  
 F:390-494/Domain: cadherin repeat homology <CR4>  
 F:497-599/Domain: cadherin repeat homology <CR5>  
 F:602-708/Domain: cadherin repeat homology <CR6>  
 F:718-822/Domain: cadherin repeat homology <CR7>  
 F:831-942/Domain: cadherin repeat homology <CR8>  
 F:948-1049/Domain: cadherin repeat homology <CR9>  
 F:1052-1153/Domain: cadherin repeat homology <CR10>  
 F:1156-1278/Domain: cadherin repeat homology <CR11>  
 F:1281-1384/Domain: cadherin repeat homology <CR12>  
 F:1387-1489/Domain: cadherin repeat homology <CR13>  
 F:1492-1601/Domain: cadherin repeat homology <CR14>  
 F:1607-1713/Domain: cadherin repeat homology <CR15>  
 F:1717-1833/Domain: cadherin repeat homology <CR16>  
 F:1826-1922/Domain: cadherin repeat homology <CR17>  
 F:1925-2027/Domain: cadherin repeat homology <CR18>  
 F:2028-2167/Domain: cadherin repeat homology <CR19>  
 F:2169-2278/Domain: cadherin repeat homology <CR20>  
 F:2281-2384/Domain: cadherin repeat homology <CR21>  
 F:2387-2491/Domain: cadherin repeat homology <CR22>  
 F:2494-2596/Domain: cadherin repeat homology <CR23>  
 F:2599-2703/Domain: cadherin repeat homology <CR24>  
 F:2707-2810/Domain: cadherin repeat homology <CR25>  
 F:2813-2913/Domain: cadherin repeat homology <CR26>  
 F:2915-3013/Domain: cadherin repeat homology <CR27>  
 F:3014-3124/Domain: cadherin repeat homology <CR28>  
 F:3127-3229/Domain: cadherin repeat homology <CR29>  
 F:3232-3334/Domain: cadherin repeat homology <CR30>  
 F:3337-3439/Domain: cadherin repeat homology <CR31>  
 F:3442-3545/Domain: cadherin repeat homology <CR32>  
 F:3548-3651/Domain: cadherin repeat homology <CR33>  
 F:3654-3756/Domain: EGF homology <EG1>  
 F:3954-4010/Domain: EGF homology <EG2>  
 F:4011-4048/Domain: EGF homology <EG3>  
 F:4056-4089/Domain: EGF homology <EG4>  
 F:4096-4127/Domain: EGF homology <EG5>  
 F:4584-4609/Domain: transmembrane #status predicted <TM>  
 F:4610-5147/Domain: intracellular #status predicted <INT>

Query Match 54.5%; Score 48; DB 1; Length 5147;  
 Best Local Similarity 71.4%; Pred. No. 30;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Oy 2 ECHIKVDVNDNP 15  
 Db 141 EVRIKVDVNDNP 154  
 RESULT 9

IGHUC5  
 Cadherin 5 precursor - human  
 N:Alternate names: 7B4 antigen; cadherin, endothelial-specific; VE-cadherin  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1993 #sequence\_revision 13-Sep-1996 #text\_change 22-Jun-1999  
 C:Accession: S43893; S24305; A43418  
 C:Brevario, F.; Cavada, U.; Corada, M.; Martin-Padura, I.; Golay, J.; Introna, M.; L.  
 submitted to the EMBL Data Library, June 1994  
 A:Description: Molecular and functional properties of VE-cadherin (7B4/cadherin-5) a  
 A:Reference number: S43893  
 A:Accession: S43893  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-784 <BRE>  
 A:Cross-references: EMBL:X79981; NID:G599833; PIDN:CA456306.1; PID:G599834  
 R:Suzuki, S.; Sano, K.; Tanikawa, H.  
 Cell Regul. 2, 261-270, 1991  
 A:Title: Diversity of the cadherin family: evidence for eight new cadherins in neuron  
 A:Reference number: S24305; MUID:91283540  
 A:Accession: S24305  
 A:Molecule type: mRNA  
 A:Residues: 5-516; 'I', 518-784 <SUZ>  
 A:Cross-references: EMBL:X59796; NID:G639976; PIDN:CA42468.1; PID:G29593  
 R:Lampugnani, M.G.; Resnati, M.; Rafteri, M.; Picot, R.; Plascane, A.; Hoven, G.; Ru  
 J. Cell Biol. 118, 1511-1522, 1992  
 A:Title: A novel endothelial-specific membrane protein is a marker of cell-cell conta  
 A:Reference number: A43418; MUID:92394977  
 A:Accession: A43418  
 A:Molecule type: protein  
 A:Residues: 48-60; 'X', 62; 'X', 64; 108-116; 'X', 118-123; 237-238; 'X', 240; 'X', 242-252; 'X', 2  
 A:Experimental source: cultured endothelial cells  
 A:Note: sequence extracted from NCBI database (NCBI:113040, NCBI:113045, NCBI:1130  
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are through  
 C:Genetics:  
 A:Gene: GDB:CDH5  
 A:Cross-references: GDB:134230; OMIM:601120  
 A:Map position: 16q22.1-16q22.1  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
 F:1-25/Domain: signal sequence #status predicted <SIG>  
 F:26-47/Domain: propeptide #status predicted <PRO>  
 F:48-784/Product: cadherin 5 #status predicted <MAT>  
 F:48-593/Domain: extracellular #status predicted <EXT>  
 F:50-151/Domain: cadherin repeat homology <CR1>  
 F:154-258/Domain: cadherin repeat homology <CR2>  
 F:261-372/Domain: cadherin repeat homology <CR3>  
 F:375-479/Domain: cadherin repeat homology <CR4>  
 F:481-587/Domain: cadherin repeat homology <CR5>  
 F:594-620/Domain: transmembrane #status predicted <TM>  
 F:621-784/Domain: intracellular #status predicted <INT>  
 F:736-753/Region: serine-rich  
 F:61,112,157,362,442,523,535/Binding site: carbohydrate (Asn) (covalent) #status pred

Query Match 54.5%; Score 48; DB 1; Length 784;  
 Best Local Similarity 81.8%; Pred. No. 46;  
 Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 Oy 5 IKVVDVNDNP 15  
 Db 139 IKVVDVNDNP 149  
 RESULT 10  
 145858  
 desmocollin - bovine  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 19-Dec-1997 #sequence\_revision 19-Dec-1997 #text\_change 20-Aug-1999  
 C:Accession: I45858  
 R:Yue, K.K.; Holton, J.L.; Clarke, J.P.; Hyam, J.L.; Hashimoto, T.; Chidgey, M.A.; Ga  
 J. Cell Sci. 108, 2163-2173, 1995  
 A:Title: Characterisation of a desmocollin isoform (bovine DSC3) exclusively expresse  
 A:Reference number: I45858; MUID:95403557

A:Accession: T45858  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-896 <YUE>  
A:Cross-references: GB:L33774; NID:g9144820; PIDN:AA041625.1; PID:g9144821  
C:Genetics:  
A:Gene: DSC3  
A:Initons: 831/3  
C:Superfamily: cadherin; cadherin repeat homology

Query Match 53.4%; Score 47; DB 2; Length 896;  
Best Local Similarity 69.2%; Pred. No. 7.6;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
OY 3 CNIKVDVNDNFP 15  
DB 340 CITVRKSDNDLP 352

RESULT 11  
S44887  
ZK112.7 protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 14-Sep-1994 #sequence\_revision 12-May-1995 #text\_change 09-Sep-1997  
C:Accession: S44887  
R:Du, Z.  
Submitted to the EMBL Data Library, May 1993  
A:Description: Sequence of the C. elegans cosmid ZK112.  
A:Reference number: S44613  
A:Accession: S44887  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-3343 <DUZ>  
A:Cross-references: EMBL:L4324; NID:g289740; PID:g289742  
C:Genetics:  
A:Initons: 39/1; 74/3; 122/3; 209/2; 287/3; 333/2; 489/1; 611/2; 714/1; 939/3; 1002/2; 1  
C:Keywords: cytoskeleton; transmembrane protein

Query Match 53.4%; Score 47; DB 2; Length 3343;  
Best Local Similarity 53.3%; Pred. No. 28;  
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;  
OY 1 CECNIRKVDVNDNFP 15  
DB 213 CHNLTILINDNFP 227

RESULT 12  
S38673  
desmoglein 2 - human  
N:Alternate names: desmoglein HDCC  
C:Species: Homo sapiens (man)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999  
C:Accession: S38673; B38872  
R:Zimbelmann, R.  
Submitted to the EMBL Data Library, September 1993  
A:Reference number: S38673  
A:Accession: S38673  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1117 <ZIM>  
A:Cross-references: EMBL:Z26317; NID:g416177; PIDN:CAA81226.1; PID:g416178  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Eur. J. Cell Biol. 55, 200-208, 1991  
A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide  
A:Reference number: A38872; MUID:92037656  
A:Accession: B38872  
A:Molecule type: mRNA  
A:Residues: 777-1117 <KOC>  
A:Cross-references: GB:S64273  
C:Genetics:

A:Gene: GDB:DSC2  
A:Cross-references: GDB:128808; OMIM:125671  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane prote  
F:51-158/Domain: cadherin repeat homology <CR1>  
F:161-271/Domain: cadherin repeat homology <CR2>

Query Match 52.3%; Score 46; DB 2; Length 1117;  
Best Local Similarity 64.3%; Pred. No. 14;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
OY 2 ECNIRKVDVNDNFP 15  
DB 143 ELRIKVIDINDNFP 156

RESULT 13  
D38992  
cadherin 8 - human  
C:Species: Homo sapiens (man)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 20-Aug-1999  
C:Accession: D38992  
R:Suzuki, S.; Sano, K.; Tanihara, H.  
Cell Regul. 2, 261-270, 1991  
A:Title: Diversity of the cadherin family: evidence for eight new cadherins in nervous  
A:Reference number: S24305; MUID:91283540  
A:Accession: D38992  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-793 <SUZ>  
A:Cross-references: GB:L34060; NID:g506411; PIDN:AA035628.1; PID:g506412  
C:Genetics:  
A:Gene: GDB:CDH8  
A:Cross-references: GDB:5822911  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication

Query Match 52.3%; Score 46; DB 2; Length 793;  
Best Local Similarity 72.7%; Pred. No. 9.7;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 5 IKYVDVNDNFP 15  
DB 148 IKYVDINDNAP 158

RESULT 14  
G02678  
cadherin-14 - human  
C:Species: Homo sapiens (man)  
C:Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999  
C:Accession: G02678  
R:Shibata, T.; Shimoyama, Y.; Gotch, M.; Hirohashi, S.  
Submitted to the EMBL Data Library, May 1996  
A:Reference number: H01584  
A:Accession: G02678  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-790 <SHI>  
A:Cross-references: EMBL:U59325; NID:g1389852; PIDN:AA02933.1; PID:g1389853  
C:Superfamily: cadherin; cadherin repeat homology

Query Match 52.3%; Score 46; DB 2; Length 790;  
Best Local Similarity 72.7%; Pred. No. 9.7;  
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
OY 5 IKYVDVNDNFP 15  
DB 147 IKYVDINDNAP 157



## RESULT 15

T09055  
Protocadherin 68 - human  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 11-Jun-1999  
C:Accession: T09055  
R:jin, P.; Xu, H.; Israel, D.  
submitted to the EMBL Data Library, October 1997  
A:Reference number: Z16540  
A:Accession: T09055  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-889 <JIN>  
A:Cross-references: EMBL:AF029343; NID:g2599501; PID:g2599502  
C:Genetics:  
A:Gene: PCH68

Query Match 52.3%; Score 46; DB 2; Length 889;  
Best Local Similarity 40.0%; Pred. No. 11;  
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
QY 1 CECNIKVDNDNFP 15  
| : : : : : |  
Db 116 CMKVEIQDINDNAP 130

Search completed: January 7, 2000, 08:52:44  
Job time: 2274 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:39 ; Search time 28.55 Seconds  
(without alignments)  
15.282 Million cell updates/sec

Title: US-08-991-628-5  
Perfect score: 88  
Sequence: 1 CECNIRKVDVNDNFP 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38:\*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	100.0	999	1 DSG3_HUMAN	P39266 homo sapien
2	74	84.1	1043	1 DSG1_BOVIN	Q03763 bos taurus
3	74	84.1	1049	1 DSG1_HUMAN	Q03763 bos taurus
4	49	55.7	863	1 DSG2_BOVIN	P33545 bos taurus
5	49	55.7	902	1 DSG2_MOUSE	P33192 mus musculu
6	48	54.5	784	1 CAD5_HUMAN	P33151 homo sapien
7	48	54.5	783	1 CAD5_MOUSE	P55284 mus musculu
8	48	54.5	5147	1 FAT_DROME	P33450 drosophila
9	47	53.4	782	1 CAD5_PIG	Q02840 sus scrofa
10	47	53.4	896	1 DSG3_BOVIN	Q20860 bos taurus
11	47	53.4	3343	1 YOG7_CAEEL	P34616 caenorhabdi
12	46	52.3	793	1 CAD8_HUMAN	P55286 homo sapien
13	46	52.3	799	1 CAD8_MOUSE	P97291 mus musculu
14	46	52.3	790	1 CADE_HUMAN	Q13634 homo sapien
15	46	52.3	1117	1 DSG2_HUMAN	Q14126 homo sapien
16	45	51.1	794	1 CAD6_HUMAN	P55289 homo sapien
17	45	51.1	827	1 CAD6_MOUSE	P55289 homo sapien
18	45	51.1	901	1 DSG2_HUMAN	Q03487 homo sapien
19	45	51.1	885	1 IF2_AOUAE	O67825 aquifex aeo
20	44	50.0	804	1 CAD1_MOUSE	P03803 mus musculu
21	44	50.0	790	1 CAD6_HUMAN	P55285 homo sapien
22	44	50.0	790	1 CAD6_MOUSE	P55285 homo sapien
23	44	50.0	790	1 CAD6_MOUSE	P55285 homo sapien
24	44	50.0	790	1 CAD6_MOUSE	P55285 homo sapien
25	43	48.9	796	1 CAD8_HUMAN	P55287 homo sapien
26	43	48.9	796	1 CAD8_MOUSE	P55288 mus musculu
27	42.5	48.9	135	1 YN8Q_YEAST	P53738 saccharomyc
28	42	47.7	789	1 CAD6_MOUSE	P79995 gallus galli
29	41	46.6	872	1 CAD1_XENLA	P30944 xenopus lae
30	41	46.6	171	1 PCRI_SCHPO	Q09926 schizosach
31	40	45.5	732	1 CAD8_MOUSE	P33145 gallus galli
32	40	45.5	712	1 CAD6_MOUSE	P33150 gallus galli
33	40	45.5	527	1 YAD7_SCHPO	Q09833 schizosach
34	39.5	44.3	1986	1 W4_EMENT	Q03149 emericella
35	39	44.3	887	1 CAD1_CHICK	P08641 gallus galli
36	39	44.3	426	1 C1SY_HELPY	P56062 helicobacte
37	39	44.3	894	1 DSG1_HUMAN	Q08554 homo sapien
38	39	44.3	896	1 DSG3_HUMAN	Q14574 homo sapien
39	39	44.3	294	1 DSG3_MOUSE	P55850 mus musculu

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	999 AA.	
DSG3_HUMAN	1				
DSG3_HUMAN	1				
AC	P39266				
DT	01-OCT-1993 (Rel. 27, Created)				
DT	01-OCT-1993 (Rel. 27, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	DESMOGLEIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).				
GN	DSG3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
OC	Eukaryota; Primates; Catarrhini; Homidae; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE: 92069753.				
RA	AMAGAI M., KLAUS-KOYTUN V., STANLEY J.R.;				
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion."				
RT	Cell 67:869-877(1991).				
RL	Cell 67:869-877(1991).				
CC	- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.				
CC	INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.				
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, ESOPHAGUS AND CARCINOMAS.				
CC	- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).				
CC	- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES AGAINST DSG3.				
CC	- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: M76482; AAA60230.1; .				
DR	PIR: A41088; ITHUG3.				
DR	HSSP: P09803; IEDH.				
DR	MIM: 169615.				
DR	PFAM: PF00028; cadherin; 4.				
DR	PROSITE: PS00232; CADHERIN; 3.				
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; Calcium-binding; Repeat.				
KW	SIGNAL				
FT	PROPEP	24	49		POTENTIAL.
FT	CHAIN	50	999		POTENTIAL.
FT	DOMAIN	50	615		DESMOGLEIN 3.
FT	TRANSMM	616	640		EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	641	999		POTENTIAL.
FT	REPEAT	159	158		CYTOLASMIC (POTENTIAL).
FT	REPEAT	159	268		CADHERIN 1.
FT	REPEAT	269	383		CADHERIN 2.
FT	REPEAT	384	499		CADHERIN 3.
FT	REPEAT	910	935		CADHERIN 4.
FT	REPEAT	936	966		DESMOGLEIN REPEAT 1.
FT	REPEAT	936	966		DESMOGLEIN REPEAT 2.
FT	CARBOHYD	110	110		POTENTIAL.

FT CARBOHYD 180 180 POTENTIAL.  
 FT CARBOHYD 459 459 POTENTIAL.  
 FT CARBOHYD 545 545 POTENTIAL.  
 SQ SEQUENCE 999 AA: 107503 MW: 489156AE CRC32:

Query Match 100.0%; Score 88; DB 1; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 8, 1e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CECNIKVKVDNDNFP 15  
 |||||  
 DB 251 CECNIKVKVDNDNFP 265

RESULT 2  
 ID DSG1\_BOVIN STANDARD; PRT; 1043 AA.  
 AC 003763;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
 GN DSG1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-MUZZLE EPITHELIUM;  
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;  
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 44-1043 FROM N.A.  
 RC TISSUE-MUZZLE EPITHELIUM;  
 RX MEDLINE: 91168965.  
 RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,  
 RA ZIMBELMANN R., FRANK W.W.;  
 RT "Identification of desmoglein, a constitutive desmosomal  
 glycoprotein, as a member of the cadherin family of cell adhesion  
 molecules."  
 RL Eur. J. Cell Biol. 53:1-12(1990).  
 RN [3]  
 RN REVISIONS, AND SEQUENCE OF 101-123.  
 RX MEDLINE: 92037656.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
 RA FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 polypeptide and identification of a second type of desmoglein gene."  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 RN [4]  
 RP SEQUENCE OF 44-493 FROM N.A.  
 RX MEDLINE: 91097553.  
 RA GOODMAN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;  
 RT "Desmoglein shows extensive homology to the cadherin family of cell  
 adhesion molecules."  
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).  
 CC - FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC - TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.  
 CC - DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC - SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC EMBL: X58466; CAA11380.1;  
 CC EMBL: X57784; CAA40930.1;  
 CC EMBL: M58165; AAA62709.1;  
 CC PIR: S14603; IJBOG1.  
 CC HSSP: P09803; IEDH.  
 CC PIR: P00028; cadherin. 3.  
 CC PROSITE: PS00322; CADHERIN; 2.  
 CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 CC Calcium-binding; Repeat.  
 CC SIGNAL 1 23  
 CC PROPEP 24 49  
 CC CHAIN 50 1043  
 CC DOMAIN 50 348  
 CC TRANSMEM 549 573  
 CC DOMAIN 574 1043  
 CC REPEAT 50 158  
 CC REPEAT 159 270  
 CC REPEAT 271 385  
 CC REPEAT 386 498  
 CC REPEAT 819 845  
 CC REPEAT 846 875  
 CC REPEAT 876 905  
 CC REPEAT 906 933  
 CC REPEAT 934 962  
 CC DOMAIN 963 1012  
 CC CARBOHYD 110 110  
 CC CARBOHYD 180 180  
 CC CARBOHYD 496 496  
 CC CONFLICT 124 124  
 CC SEQUENCE 1043 AA: 112243 MW: 13898584 CRC32:  
 POTENTIAL.  
 POTENTIAL.  
 I -> V (IN REF. 4).  
 POTENTIAL.

Query Match 84.1%; Score 74; DB 1; Length 1043;  
 Best Local Similarity 80.0%; Pred. No. 0.00016;  
 Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 CECNIKVKVDNDNFP 15  
 |||||  
 DB 253 CECNIKVKVDNDNFP 267

RESULT 3  
 ID DSG1\_HUMAN STANDARD; PRT; 1049 AA.  
 AC 002413;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
 GN DSG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE-KERATINOCYTES;  
 RX MEDLINE: 91271279.  
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALOTIS P., POYNTER D.,  
 RA ARNEJAN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,  
 RA BUXTON R.S., MAGEE A.I.;  
 RT "Desmosomal glycoprotein Dgl, a component of intercellular desmosome  
 junctions, is related to the cadherin family of cell adhesion  
 molecules."  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).  
 CC - FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC - TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSILLAD ESOPHAGUS.  
 CC - DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: X56654; CAA39976.1; -  
 CC DR PIR: S16906; IJHGH1.  
 CC DR HSSP: P09803; 1EDH.  
 CC DR MIM: 125670; -  
 CC DR PIR: P00028; cadherin; 4.  
 CC DR PROSITE: PS00232; CADHERIN; 2.  
 CC KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 CC Calcium-binding; Repeat.  
 CC FT SIGNAL 1 23 POTENTIAL.  
 CC FT PROPEP 24 49 POTENTIAL.  
 CC FT CHAIN 50 1049 DESMOGLEIN 1.  
 CC FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 546 570 POTENTIAL.  
 CC FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).  
 CC FT REPEAT 50 158 CADHERIN 1.  
 CC FT REPEAT 159 270 CADHERIN 2.  
 CC FT REPEAT 271 385 CADHERIN 3.  
 CC FT REPEAT 386 497 CADHERIN 4.  
 CC FT REPEAT 498 669 DESMOGLEIN REPEAT 1.  
 CC FT REPEAT 670 839 DESMOGLEIN REPEAT 2.  
 CC FT REPEAT 840 899 DESMOGLEIN REPEAT 3.  
 CC FT REPEAT 900 927 DESMOGLEIN REPEAT 4.  
 CC FT REPEAT 928 956 DESMOGLEIN REPEAT 5.  
 CC FT DOMAIN 969 1019 GLY/SER-RICH.  
 CC FT CARBOHYD 36 36 POTENTIAL.  
 CC FT CARBOHYD 110 110 POTENTIAL.  
 CC FT CARBOHYD 180 180 POTENTIAL.  
 CC FT SEQUENCE 1049 AA; 113715 MW; FDD/9961 CRC32;  
 CC -----  
 CC Query Match Score 74; DB 1; Length 1049;  
 CC Best Local Similarity 80.0%; Pred. No. 0.00016;  
 CC Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 CC -----  
 CC QY 1 CECNIKVDNDNFP 15  
 CC DB 253 CECNIKILDVNDNIP 267  
 CC -----  
 CC RESULT 4  
 CC DSC2\_BOVIN STANDARD: PRT; 863 AA.  
 CC AC P33545;  
 CC DT 01-FEB-1994 (Rel. 28, Created)  
 CC DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 CC DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 CC DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN)  
 CC DE (FRAGMENT).  
 CC GN DSC2.  
 CC OS Bos taurus (Bovine).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 CC OC Bovinae; Bos.  
 CC RN (1)  
 CC RP SEQUENCE FROM N.A.  
 CC RC TISSUE-MUZZLE EPITHELIUM;  
 CC RX MEDLINE: 92108053.  
 CC RA KOCH P.J., GOLDSCHMIDT M.D., ZIMMELMANN R., TROYANOVSKY R.,  
 CC RA FRAKE W.W.;  
 CC "Complexity and expression patterns of the desmosomal cadherins";  
 CC RT Proc. Natl. Acad. Sci. U.S.A. 89:353-357(1992).  
 CC CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED  
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS

CC MEDIANING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL  
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL  
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: FORMS 2A (SHOWN HERE) AND 2B ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE SAME GENE.  
 CC -1- TISSUE SPECIFICITY: ESOPHAGUS AND RUMEN. WEAKLY IN EPITHELIA AND  
 CC CARDIAC MUSCLE.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -----  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL: M81190; AAA30782.1; -  
 CC DR EMBL: M81190; AAA30783.1; -  
 CC DR PIR: A41799; IJBODC.  
 CC DR PIR: B41799; IJBODC.  
 CC DR HSSP: P09803; 1SRH.  
 CC DR PIR: P00028; cadherin; 5.  
 CC DR PROSITE: PS00232; CADHERIN; 3.  
 CC KW Cell adhesion; Glycoprotein; Transmembrane; Repeat;  
 CC Alternative splicing; Cytoskeleton; Calcium-binding.  
 CC FT PROPEP 1 89 POTENTIAL.  
 CC FT CHAIN 90 863 DESMOCOLLIN 2A/2B.  
 CC FT DOMAIN 90 644 EXTRACELLULAR (POTENTIAL).  
 CC FT TRANSMEM 645 665 POTENTIAL.  
 CC FT DOMAIN 666 863 CYTOPLASMIC (POTENTIAL).  
 CC FT REPEAT 90 197 CADHERIN 1.  
 CC FT REPEAT 198 309 CADHERIN 2.  
 CC FT REPEAT 310 423 CADHERIN 3.  
 CC FT REPEAT 424 528 CADHERIN 4.  
 CC FT REPEAT 529 644 CADHERIN 5.  
 CC FT CARBOHYD 120 120 POTENTIAL.  
 CC FT CARBOHYD 346 346 POTENTIAL.  
 CC FT CARBOHYD 495 495 POTENTIAL.  
 CC FT CARBOHYD 579 579 POTENTIAL.  
 CC FT VARIANT 264 264 K -> Q.  
 CC FT VARIANT 333 333 R -> Q.  
 CC FT VARSPLIC 799 809 KVOGDODDNTH -> ESIRGHTLVKN (IN FORM 2B).  
 CC FT VARSPLIC 810 863 MISSING (IN FORM 2B).  
 CC FT SEQUENCE 863 AA; 95874 MW; 2173F06E CRC32;  
 CC -----  
 CC Query Match Score 49; DB 1; Length 863;  
 CC Best Local Similarity 61.5%; Pred. No. 1.6;  
 CC Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 CC -----  
 CC QY 3 CNIKYVDNDNFP 15  
 CC DB 295 CIINIEDVDNDNFP 307  
 CC -----  
 CC RESULT 5  
 CC DSC2\_MOUSE STANDARD: PRT; 902 AA.  
 CC AC P55292; Q64734;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 CC DE DESMOCOLLIN 2A/2B PRECURSOR (EPITHELIAL TYPE 2 DESMOCOLLIN).  
 CC DE DSC2 OR DSC3.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.  
 CC RN [1]



```

DR HSSP: P09803; 1EDH.
DR MIM: 601130; -.
DR PFAM: PF00028; cadherin. 5.
DR PFAM: PF01049; Cadherin_C-term. 1.
DR PROSITE: PS00232; CADHERIN. 3.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat; Signal.
FT SIGNAL 1 25
FT PROPEP 26 47
FT CHAIN 48 784
FT DOMAIN 48 593
FT TRANSMEM 594 620
FT DOMAIN 621 784
FT REPEAT 48 151
FT REPEAT 152 258
FT REPEAT 259 372
FT REPEAT 373 477
FT REPEAT 478 593
FT DOMAIN 736 753
FT CARBOHYD 61 61
FT CARBOHYD 112 112
FT CARBOHYD 157 157
FT CARBOHYD 362 362
FT CARBOHYD 442 442
FT CARBOHYD 523 523
FT CARBOHYD 535 535
FT CONFLICT 517 517
SQ SEQUENCE 784 AA; 87528 MM; C2C5CD71 CRC32;

Query Match
Best Local Similarity 81.8%; Score 48; DB 1; Length 784;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 5 IKVKDNDNFP 15
DB 139 IKVHDNDNWP 149

RESULT 7
CAD5_MOUSE STANDARD; PRT; 783 AA.
ID CAD5_MOUSE
AC P55284;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE VASCULAR ENDOTHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).
GN CDH5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-BRAIN CAPILLARY;
RX MEDLINE: 96141083.
RA BREIER G., BREVIARIO F., CAVEDA L., BERTHIER R., SCHNURCH H.,
RA GORSCH U., VESTWEBER D., RISAU W., DEJANA E.,
RT "Molecular cloning and expression of murine vascular endothelial-
RT cadherin in early stage development of cardiovascular system.";
RL Blood 87:630-641(1996).
BI -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A
CC IMPORTANT ROLE IN ENDOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE
CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. NERVOUS
CC SYSTEM.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X83930; CAA58782.1; -.
DR HSSP: P09803; 1EDH.
DR MGD: MGI:105057; CDH5.
DR PFAM: PF00028; cadherin. 5.
DR PFAM: PF01049; Cadherin_C-term. 1.
DR PROSITE: PS00232; CADHERIN. 3.
KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;
KW Calcium-binding; Repeat; Signal.
FT SIGNAL 1 25
FT PROPEP 26 47
FT CHAIN 48 783
FT DOMAIN 48 592
FT TRANSMEM 593 619
FT DOMAIN 620 783
FT REPEAT 46 148
FT REPEAT 149 255
FT REPEAT 256 370
FT REPEAT 371 475
FT REPEAT 476 592
FT DOMAIN 737 752
FT CARBOHYD 59 59
FT CARBOHYD 154 154
FT CARBOHYD 440 440
FT CARBOHYD 522 522
FT CARBOHYD 534 534
SQ SEQUENCE 783 AA; 87847 MM; DB071215 CRC32;

Query Match
Best Local Similarity 63.6%; Score 48; DB 1; Length 783;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 5 IKVKDNDNFP 15
DB 243 IRLDINDNFP 253

RESULT 8
FAT_DROME STANDARD; PRT; 5147 AA.
ID FAT_DROME
AC P33750;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).
GN FAT.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92069752.
RA MAHONEY P.A., WEBER U., ONOFRECHUK P., BIESSMANN H., BRYANT P.J.,
RA GOODMAN C.S.;
RT "The fat tumor suppressor gene in Drosophila encodes a novel member
RT of the cadherin gene superfamily.";
RL Cell 67:853-868(1991).
BI -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC,
CC TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN
CC DIFFERENTIATION AND MORPHOGENESIS, AND DEATH DURING THE PUPAL
CC STAGE.
CC -1- MISCELLANEOUS: SIMILARITY: CONTAINS 37 CADHERINS-TYPE REPEATS.
CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.
CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

```

Query Match	Best Local Similarity	Score	DB	Length
2 ECNKKVNDNDNFP 15	54.53;	48;	1;	5147;
10; Conservative	71.48;	Pred. NO. 14;		
Matches	10;	Mismatches	4;	Indels 0; Gaps 0
DB 141 EVRIKVLVDNDNSP 154				
RESULT 9				
CAD5_PIG	STANDARD;	PRT;	782 AA.	
AC 002840;				
DT 15-JUL-1998 (Rel. 36, Created)				
DT 15-JUL-1998 (Rel. 36, Last sequence update)				
DT 15-JUL-1998 (Rel. 36, Last annotation update)				
DE VASCUTAR ENDOHELIAL-CADHERIN PRECURSOR (VE-CADHERIN) (CADHERIN-5).				
GN CDH5.				
OS Sus scrofa (Pig);				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
CC Eutheria; Cetartiodactyla; Suidae; Sus.				
NP [1]				
SEQUENCE FROM N.A.				

RA KILSHAW P.J.:  
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY REPRESENTATIONALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THIS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. THIS CADHERIN MAY PLAY A  
 CC IMPORTANT ROLE IN ENOTHELIAL CELL BIOLOGY THROUGH CONTROL OF THE  
 CC COHESION AND ORGANIZATION OF THE INTERCELLULAR JUNCTIONS. IT  
 CC ASSOCIATES WITH ALPHA-CATENIN FORMING A LINK TO THE CYTOSKELETON  
 CC (BY SIMILARITY).  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. FOUND AT CELL-CELL  
 CC BOUNDARIES AND PROBABLY AT CELL-MATRIX BOUNDARIES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y13919; CAA74225.1; -  
 DR HSSP: P09803; 1EDH.  
 DR PFAM: PF00028; cadherin; 5.  
 DR PFAM: PF01049; Cadherin\_C-term; 1.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KM Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KW Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 44 POTENTIAL.  
 FT CHAIN 45 782 VASCULAR ENDOTHELIAL-CADHERIN.  
 FT DOMAIN 45 592 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 593 619 POTENTIAL.  
 FT DOMAIN 620 782 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 45 148 CADHERIN 1.  
 FT REPEAT 149 255 CADHERIN 2.  
 FT REPEAT 256 370 CADHERIN 3.  
 FT REPEAT 371 475 CADHERIN 4.  
 FT REPEAT 476 592 CADHERIN 5.  
 FT DOMAIN 593 751 SER-RICH.  
 FT CARBOHYD 58 58 POTENTIAL.  
 FT CARBOHYD 154 154 POTENTIAL.  
 FT CARBOHYD 360 360 POTENTIAL.  
 FT CARBOHYD 440 440 POTENTIAL.  
 FT CARBOHYD 522 522 POTENTIAL.  
 FT CARBOHYD 534 534 POTENTIAL.  
 SQ SEQUENCE 782 AA; 87546 MW; 007F70E0 CRC32;

Query Match 53.4%; Score 47; DB 1; Length 782;  
 Best Local Similarity 72.7%; Pred. No. 3.1;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 IKYKVDNDNF 15  
 DB 136 IKVHDINDNP 146  
 RESULT 10  
 DSC3\_BOVIN STANDARD; PRT; 896 AA.  
 AC Q28060; Q28061; Q28176;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOCOLLIN 3A/3B PRECURSOR.  
 GN DSC3.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Ruminantia; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95403557.  
 RA YUE K.K.M., HOLTON J.L., CLARKE J.P., HYAM J.L.M., HASHIMOTO T.,  
 RA CHIDEX M.A.J., GARROD D.R.;  
 RT "Characterisation of a desmocollin isoform (bovine DSC3) exclusively  
 RT expressed in lower layers of stratified epithelia."  
 RL J. Cell Sci. 108:2163-2173(1995).  
 RP [2]  
 RP SEQUENCE OF 686-814 FROM N.A.  
 RC TISSUE-EPIDERMIS;  
 RX MEDLINE: 94308280.  
 RA LEGAN P.K., YUE K.K.M., CHIDEX M.A.J., HOLTON J.L., WILKINSON R.W.,  
 RA GARROD D.R.;  
 RT "The bovine desmocollin family: a new gene and expression patterns  
 RT reflecting epithelial cell proliferation and differentiation."  
 RL J. Cell Biol. 126:507-518(1994).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS. INVOLVED  
 CC IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS  
 CC MEDIATING CELL-CELL ADHESION. MAY CONTRIBUTE TO EPIDERMAL CELL  
 CC POSITIONING (STRATIFICATION) BY MEDIATING DIFFERENTIAL  
 CC ADHESIVENESS BETWEEN CELLS THAT EXPRESS DIFFERENT ISOFORMS.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: FORMS 3A AND 3B ARE PRODUCED BY ALTERNATIVE  
 CC SPLICING OF THE SAME GENE.  
 CC -1- TISSUE SPECIFICITY: STRATIFIED EPITHELIA ONLY (EPIDERMIS, TONGUE,  
 CC ESOPHAGUS AND ROMEN).  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: L33774; AAC41625.1; -  
 DR EMBL: L33774; AAC41626.1; -  
 DR EMBL: X75783; CAA53427.1; -  
 DR HSSP: P09803; 1SDH.  
 DR PFAM: PF00028; cadherin; 5.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KM Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;  
 KW Alternative splicing; Cytoskeleton; Calcium-binding.  
 FT SIGNAL 1 26  
 FT PROPEP 27 134 POTENTIAL.  
 FT CHAIN 135 896 DESMOCOLLIN 3A/3B.  
 FT DOMAIN 135 690 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 691 711 POTENTIAL.  
 FT DOMAIN 712 896 CYTOPLASMIC (POTENTIAL).  
 FT REPEAT 135 242 CADHERIN 1.  
 FT REPEAT 243 354 CADHERIN 2.  
 FT REPEAT 355 471 CADHERIN 3.  
 FT REPEAT 472 579 CADHERIN 4.  
 FT DOMAIN 580 690 CADHERIN 5.  
 FT VARSPLIC 832 839 KLICND -> ESIRGHTG (IN FORM 3B).  
 FT VARSPLIC 840 896 MISSING (IN FORM 3B).  
 FT CARBOHYD 165 165 POTENTIAL.  
 FT CARBOHYD 391 391 POTENTIAL.  
 FT CARBOHYD 346 346 POTENTIAL.  
 FT CARBOHYD 629 629 POTENTIAL.  
 FT CARBOHYD 686 687 POTENTIAL.  
 SQ SEQUENCE 896 AA; 99687 MW; E5668408 CRC32;

Query Match 53.4%; Score 47; DB 1; Length 896;  
 Best Local Similarity 69.2%; Pred. No. 3.5;  
 Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 3 CNIKYKVDNDNF 15  
 DB 136 IKVHDINDNP 146



Db		340	CITIVKNSNDNLP	352
RESULT	11			
ID	YOG7_CAEEL	STANDARD;	PRT; 3343 AA.	
AC	P34616:			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	HYPOTHEETICAL 375.7 KD PROTEIN ZK112.7 IN CHROMOSOME III PRECURSOR.			
GN	Zk112.7.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditidae;			
OC	Rhabditiina; Rhabditoidea; Rhabditiidae; Pelodetinae; Caenorhabditis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RX	MEDLINE: 94150718.			
RA	WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BOFFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A., CRAXTON N., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A., FUTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIR M., JOHNSON L., JONES M., KERSHAN J., KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SIMS N., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R., SULTON J., THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLIDAN P.;			
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."			
RL	Nature 368:32-38(1994).			
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (POTENTIAL).			
CC	-1- SIMILIARY: BELONGS TO THE CADHERIN FAMILY. STRONG, TO D.MELANOCASTER FAT TUMOR SUPPRESSOR.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: L14324; AAA28182.1; .			
DR	PIR: S44887; S44887.			
DR	MORPEP: ZK112.7; CE00378.			
DR	PFAM: PF00028; cadherin_11.			
DR	PROSITE: PS00232; CADHERIN_8.			
KM	PROSITE: PS00186; EGF_2; UNKNOWN_1.			
KW	Hypothetical protein; Cell adhesion; Signal; Transmembrane; Glycosylation; Glycoprotein; Calcium-binding; Repeat.			
FT	SIGNAL	1	?	POTENTIAL.
FT	CHAIN	?	3343	HYPOTHETICAL PROTEIN ZK112.7.
FT	DOMAIN	?	3228	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	3229	3250	POTENTIAL.
FT	DOMAIN	3251	3343	CYTOSOLIC (POTENTIAL).
FT	CARBOHYD	22	.22	POTENTIAL.
FT	CARBOHYD	149	149	POTENTIAL.
FT	CARBOHYD	250	250	POTENTIAL.
FT	CARBOHYD	288	288	POTENTIAL.
FT	CARBOHYD	369	369	POTENTIAL.
FT	CARBOHYD	467	467	POTENTIAL.
FT	CARBOHYD	612	612	POTENTIAL.
FT	CARBOHYD	752	752	POTENTIAL.
FT	CARBOHYD	806	806	POTENTIAL.
FT	CARBOHYD	941	941	POTENTIAL.
FT	CARBOHYD	966	966	POTENTIAL.
FT	CARBOHYD	970	970	POTENTIAL.
FT	CARBOHYD	985	985	POTENTIAL.
FT	CARBOHYD	1042	1042	POTENTIAL.
FT	CARBOHYD	1335	1335	POTENTIAL.

[illegible]

CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN BRAIN. FOUND IN CERTAIN  
 CC NERVE CELL LINES, SUCH AS RETINOBLASTS, GLIOMA CELLS AND  
 CC NEUROBLASTS.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L34060; AAA35628.1; -  
 DR HSSP: P15116; INCI.  
 DR MIM: 603008; -  
 DR PFAM: PF00028; cadherin. 5.  
 DR PFAM: PF01049; Cadherin\_C-term. 1.  
 DR PROSITE: PS00232; CADHERIN. 3.  
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KM Calcium-binding; Repeat; Signal.  
 FT SIGNAL. 1  
 FT PROPEP 54  
 FT CHAIN ? 54  
 FT DOMAIN 55 793  
 FT TRANSMEM 55 614  
 FT DOMAIN 615 635  
 FT TRANSMEM 615 635  
 FT DOMAIN 636 793  
 FT REPEAT 55 160  
 FT REPEAT 161 269  
 FT REPEAT 270 384  
 FT REPEAT 385 487  
 FT REPEAT 488 609  
 FT CARBOHYD 26 26  
 FT CARBOHYD 50 50  
 FT CARBOHYD 181 181  
 FT CARBOHYD 456 456  
 FT CARBOHYD 466 466  
 FT CARBOHYD 537 537  
 FT SEQUENCE 793 AA; 87570 MW; 5477875A CRC32;  
 SQ

Query Match  
 Best Local Similarity 52.3%; Score 46; DB 1; Length 793;  
 Pred. No. 4.5;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 IKYKVDNDFP 15  
 DB 148 IKYQDINDNAP 158

RESULT 13  
 CAD8\_MOUSE STANDARD; PRT; 799 AA.  
 AC P97291;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE CADHERIN-8 PRECURSOR.  
 GN CDH8.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SWISS WEBSTER / NIH;  
 RX MEDLINE: 97174321.  
 RA KOREMATSU K., REDIES C.;  
 RT "Restricted expression of cadherin-8 in segmental and functional  
 RT subdivisions of the embryonic mouse brain.";

RL Dev. Dyn. 208:178-189(1997).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X95600; CAA64857.1; -  
 DR HSSP: P15116; INCI.  
 DR MGD: MGI:107434; CDH8.  
 DR PFAM: PF00028; cadherin. 5.  
 DR PFAM: PF01049; Cadherin\_C-term. 1.  
 DR PROSITE: PS00232; CADHERIN. 3.  
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KM Calcium-binding; Repeat; Signal.  
 FT SIGNAL. 1  
 FT PROPEP 30 61  
 FT CHAIN 62 799  
 FT DOMAIN 62 621  
 FT TRANSMEM 622 642  
 FT DOMAIN 643 799  
 FT TRANSMEM 643 799  
 FT DOMAIN 643 799  
 FT REPEAT 62 167  
 FT REPEAT 168 276  
 FT REPEAT 277 391  
 FT REPEAT 392 494  
 FT REPEAT 495 616  
 FT CARBOHYD 188 188  
 FT CARBOHYD 463 463  
 FT CARBOHYD 473 473  
 FT CARBOHYD 544 544  
 FT SEQUENCE 799 AA; 88200 MW; F1B2625D CRC32;  
 SQ

Query Match  
 Best Local Similarity 52.3%; Score 46; DB 1; Length 799;  
 Pred. No. 4.5;  
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 5 IKYKVDNDFP 15  
 DB 155 IKYQDINDNAP 165

RESULT 14  
 CADE\_HUMAN STANDARD; PRT; 790 AA.  
 AC Q13634;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CADHERIN-14 PRECURSOR.  
 GN CDH14.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 97184182.  
 RA SHIBATA T., SHIMOYAMA Y., GOTOH M., HIROHASHI S.;  
 RT "Identification of human cadherin-14, a novel neurally specific type  
 RT II cadherin, by protein interaction cloning.";  
 RL J. Biol. Chem. 272:5236-5240(1997).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC

MANNER IN CONNECTING CELLS. CADHERINS MAY THUS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: U59325; AAB02933.1; -.

CC HSSP: P15116; INCI.

DR HSSP: PF00028; Cadherin; 5.

DR PFAM: PF01049; Cadherin; C term; 1.

DR PROSITE: PS00232; CADHERIN; 3.

KM Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane; Calcium-binding; Repeat; Signal.

KW

FT SIGNAL 1 24

FT PROPEP 25 53 POTENTIAL.

FT CHAIN 54 790 CADHERIN-14.

FT DOMAIN 54 608 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 609 636 POTENTIAL.

FT DOMAIN 637 790 CYTOPLASMIC (POTENTIAL).

FT REPEAT 54 159 CADHERIN 1.

FT REPEAT 160 268 CADHERIN 2.

FT REPEAT 269 383 CADHERIN 3.

FT REPEAT 384 486 CADHERIN 4.

FT REPEAT 487 608 CADHERIN 5.

FT CARBOHYD 36 36 POTENTIAL.

FT CARBOHYD 255 255 POTENTIAL.

FT CARBOHYD 455 455 POTENTIAL.

FT CARBOHYD 536 536 POTENTIAL.

SO SEQUENCE 790 AA; 88072 MW; 4596258F CRC32;

Query Match

Best Local Similarity 52.3%; Score 46; DB 1; Length 790;

Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 IKVKVDNDNP 15

DB 147 IKVKVDNDNP 157

RESULT 15

DSG2\_HUMAN

ID DSG2\_HUMAN STANDARD; PRT; 1117 AA.

AC Q14126;

DT 01-NOV-1997 (Rel. 35; Created)

DT 01-NOV-1997 (Rel. 35; Last sequence update)

DE 01-NOV-1997 (Rel. 35; Last annotation update)

DE DESMOGLEIN 2 PRECURSOR (HDCG).

GN DSG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-COLON CARCINOMA;

RX MEDLINE: 94192736.

RA SCHAEFER S., KOCH P.J., FRANK W.W.;

RT "Identification of the ubiquitous human desmoglein, Dsg2, and the expression catalogue of the desmoglein subfamily of desmosomal cadherins."

RT Exp. Cell Res. 211:391-399(1994).

RL (2)

RN

RP SEQUENCE OF 777-1117 FROM N.A.

RX MEDLINE: 92037656.

RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R., FRANK W.W.;

RT "Complete amino acid sequence of the epidermal desmoglein precursor

RT polypeptide and identification of a second type of desmoglein gene.";

RL Eur. J. Cell Biol. 55:200-208(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL: Z26317; CAAB1226.1; -.

CC HSSP: P15116; INCI.

DR MIM: 125671; -.

DR PFAM: PF00028; Cadherin; 4.

DR PROSITE: PS00232; CADHERIN; 3.

KM Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal; Cytoskeleton; Calcium-binding.

KW

FT SIGNAL 1 23

FT PROPEP 24 48 POTENTIAL.

FT CHAIN 49 1117 DESMOGLEIN 2.

FT DOMAIN 49 608 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 609 633 POTENTIAL.

FT DOMAIN 634 1117 CYTOPLASMIC (POTENTIAL).

FT REPEAT 49 159 CADHERIN 1.

FT REPEAT 160 272 CADHERIN 2.

FT REPEAT 273 387 CADHERIN 3.

FT REPEAT 388 502 CADHERIN 4.

FT REPEAT 503 880 DESMOGLEIN REPEAT 1.

FT REPEAT 880 912 DESMOGLEIN REPEAT 2.

FT REPEAT 912 941 DESMOGLEIN REPEAT 3.

FT REPEAT 942 967 DESMOGLEIN REPEAT 4.

FT REPEAT 968 991 DESMOGLEIN REPEAT 5.

FT REPEAT 992 1020 DESMOGLEIN REPEAT 6.

FT REPEAT 1021 1050 DESMOGLEIN REPEAT 6.

FT CARBOHYD 111 111 POTENTIAL.

FT CARBOHYD 181 181 POTENTIAL.

FT CARBOHYD 308 308 POTENTIAL.

FT CARBOHYD 461 461 POTENTIAL.

FT CARBOHYD 513 513 POTENTIAL.

SO SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match

Best Local Similarity 52.3%; Score 46; DB 1; Length 1117;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 ECKIKVDNDNP 15

DB 143 ECKIKVDNDNP 156

Search completed: January 7, 2000, 13:25:40

Job time: 392 sec

Sat Jan 15 11:45:04 2000

us-08-991-628-5.rsp

---

Page 11

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:22 ; Search time 59.21 Seconds

(without alignments)  
15,518 Million cell updates/sec

Title: US-08-991-628-5

Perfect score: 88

Sequence: 1 CECNIKVKDVNDNFP 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTRMBL\_11:\*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	82	93.2	993	11	035902	035902 mus musculus
2	57	64.8	519	13	090425	090425 brachydanio
3	52	59.1	832	4	015336	015336 homo sapien
4	52	59.1	832	4	012864	012864 homo sapien
5	51	58.0	2163	5	001912	001912 caenorhabdi
6	50	56.8	1507	5	024298	024298 drosophila
7	50	56.8	1180	11	055134	055134 mus musculus
8	49	55.7	222	3	013737	013737 schizosacch
9	49	55.7	810	4	075279	075279 homo sapien
10	49	55.7	807	4	075288	075288 homo sapien
11	49	55.7	936	4	095206	095206 homo sapien
12	49	55.7	915	5	062328	062328 caenorhabdi
13	48	54.5	824	4	075287	075287 homo sapien
14	48	54.5	784	11	035542	035542 mus musculus
15	48	54.5	947	11	088689	088689 mus musculus
16	47	53.4	814	6	077704	077704 canis famill
17	47	53.4	889	11	088193	088193 mus musculus
18	46	52.3	889	4	014917	014917 homo sapien
19	46	52.3	4590	4	014517	014517 homo sapien
20	46	52.3	816	4	075284	075284 homo sapien
21	46	52.3	799	11	054800	054800 rattus norv
22	46	52.3	532	11	054801	054801 rattus norv
23	45	51.1	934	4	060622	060622 homo sapien
24	45	51.1	134	4	015202	015202 homo sapien
25	45	51.1	904	4	008192	008192 homo sapien

26	45	51.1	814	4	075281	075281 homo sapien
27	45	51.1	824	4	075286	075286 homo sapien
28	45	51.1	829	4	075309	075309 homo sapien
29	45	51.1	3380	5	024292	024292 drosophila
30	45	51.1	578	11	063423	063423 rattus norv
31	45	51.1	895	11	055122	055122 mus musculus
32	45	51.1	932	13	013129	013129 gallus gall
33	45	51.1	148	13	093300	093300 brachydanio
34	45	51.1	979	13	093509	093509 xenopus lae
35	44	50.0	235	2	067056	067056 aquilex aeo
36	44	50.0	1026	4	008174	008174 homo sapien
37	44	50.0	842	4	015053	015053 homo sapien
38	44	50.0	807	4	075277	075277 homo sapien
39	44	50.0	792	4	075278	075278 homo sapien
40	44	50.0	844	4	075280	075280 homo sapien
41	44	50.0	789	4	075282	075282 homo sapien
42	44	50.0	620	5	001909	001909 caenorhabdi
43	44	50.0	2809	5	061230	061230 lytechinus
44	44	50.0	3097	5	015943	015943 drosophila
45	44	50.0	829	6	028634	028634 oryctolagus

## ALIGNMENTS

RESULT 1

ID 035902 PRELIMINARY; PRT; 993 AA.

AC 035902;

DT 01-JAN-1998 (TREMBLrel. 05. Created)

DT 01-JAN-1998 (TREMBLrel. 05. Last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10. Last annotation update)

DE DESMOGLEIN 3. (FRAGMENT).

GN DSG3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BALB/C;

RA ISHIKAWA H., LI K., UETOMO J.;

RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL: U86016; AAB65091.1; -

DR PFWA: PF00028; cadherin; 4.

DR PROSITE: PS00232; CADHERIN; 2.

KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.

FT NON-TER 993

SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 93.2%; Score 82; DB 11; Length 993;

Best Local Similarity 86.7%; Pred. No. 3.2e-05;

Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 CECNIKVKDVNDNFP 15

DB 251 CECNIKVKDVNDNFP 265

RESULT 2

ID 090425 PRELIMINARY; PRT; 519 AA.

AC 090425;

DT 01-NOV-1996 (TREMBLrel. 01. Created)

DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10. Last annotation update)

DE VENTRAL NEURAL CADHERIN (FRAGMENT).

GN VNC.

OS Brachydanio rerio (Zebrafish) (Zebra danio).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Actinopterygii;

OC Neopterygii; Teleostei; Osteichthys; Cypriniformes;

OC Cyprinidae; Cyprinidae; Rasbora; Danio.

[1]  
 RN SEQUENCE FROM N.A.  
 RP FRANKLIN J.L., SARGENT T.D.;  
 RL Dev. Dyn. 206:0-0(0).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: U01419; AAB47406.1; -.  
 DR PFAM: PF00028; cadherin\_3.  
 DR PROSITE: PS00232; CADHERIN\_1.  
 DR ZFIN: ZDB-GENE-980526-170; vnc.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 FT NON TER  
 SQ SEQUENCE 519 AA; 57807 MW; E6DA0079 CRC32;

Query Match 64.8%; Score 57; DB 13; Length 519;  
 Best Local Similarity 91.7%; Pred. No. 0.16;  
 Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 4 NIKVKVDNDNFP 15  
 DB 202 NIKVKVDNDNAP 213

RESULT 3  
 ID 015336 PRELIMINARY; PRT; 832 AA.  
 AC 015336;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)  
 DE LI-CADHERIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA BOETTNER A., KRETT B., FIEGER C., DLOUHY B., BERNDOERF D.,  
 RA GOESSNER R., TAUBER R.;  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: X83228; CA58231.1; -.  
 DR PFAM: PF00028; cadherin\_7.  
 DR PROSITE: PS00232; CADHERIN\_3.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 832 AA; 92207 MW; 65C92D4C CRC32;

Query Match 59.1%; Score 52; DB 4; Length 832;  
 Best Local Similarity 64.3%; Pred. No. 1.6;  
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 ECHIKVNDNFP 15  
 DB 325 ECHIKVNDNMP 338

RESULT 4  
 ID 012864 PRELIMINARY; PRT; 832 AA.  
 AC 012864;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)  
 DE INTERINAL PEPTIDE-ASSOCIATED TRANSPORTER HPT-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE-COLON;  
 RX MEDLINE: 9420463.  
 RA DANTZIG A.H., HOSKINS J., TABAS L.B., BRIGHT S., SHEPARD R.L.,

RA JENKINS I.L., DUCKWORTH D.C., SPORTSMAN R., MACKENSEN D.,  
 RA ROSTECK P.R., SKATROD P.L.;  
 RT "Association of intestinal peptide transport with a protein related  
 to the cadherin superfamily."  
 RL Science 264:430-433(1994).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: U07969; AAA19021.1; -.  
 DR PFAM: PF00028; cadherin\_7.  
 DR PROSITE: PS00232; CADHERIN\_3.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 832 AA; 92147 MW; B5193483 CRC32;

Query Match 59.1%; Score 52; DB 4; Length 832;  
 Best Local Similarity 64.3%; Pred. No. 1.6;  
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 ECHIKVNDNFP 15  
 DB 325 ECHIKVNDNMP 338

RESULT 5  
 ID 001912 PRELIMINARY; PRT; 2163 AA.  
 AC 001912;  
 DT 01-JUL-1997 (TEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)  
 DE SIMILARITY TO MULTIPLE CADHERIN-TYPE REPEATS.  
 GN R10F2.1.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;  
 OC Rhabditina; Rhabditidae; Rhabditidae; Pelodidae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCINTYRE A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STURTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUGHAN K., WATSON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans."  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA DU Z., GATTUNG S.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA WATSON R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AF003386; AAB54266.1; -.  
 DR PFAM: PF00001; 7tm\_1; 1.  
 DR PFAM: PF00028; cadherin\_15.  
 DR PROSITE: PS00232; CADHERIN\_8.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 2163 AA; 238609 MW; 6C856652 CRC32;

Query Match 58.0%; Score 51; DB 5; Length 2163;  
 Best Local Similarity 53.3%; Pred. No. 5.9;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Oy	1	CENCKVKDVNDNFP	15
	1:::	1	
Db	1629	CKCHYIVDENDNLF	1643
RESULT	6		
ID	024298	PRELIMINARY;	PRT; 1507 AA.
AC	Q24298;		
DT	01-NOV-1996	(TREMBLrel. 01. Created)	
DT	01-NOV-1996	(TREMBLrel. 01. last sequence update)	
DT	01-MAY-1999	(TREMBLrel. 10. last annotation update)	
DE	DE-CADHERIN PRECURSOR (SHOTGUN) (SHG).		
DN	SHG OR GP150 OR DE-CADHERIN.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha		
NC	Ephydroidea; Drosophilidae; Drosophila.		
RN	[1]		
RP	SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.		
RC	TISSUE-EMBRYO.		
RX	MEDLINE; 95046887.		
RA	ODA H., UEMURA T., HARADA Y., IMAI Y., TAKICHI M.;		
RT	"A Drosophila homolog of cadherin associated with armadillo and essential for embryonic cell-cell adhesion."		
RL	Dev. Biol. 165:716-726(1994).		
CC	-1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS		
CC	THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC		
CC	MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE		
CC	SORTING OF HETEROGENEOUS CELL TYPES.		
CC	N-CADHERIN MAY BE INVOLVED IN NEURONAL RECOGNITION MECHANISM.		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.		
CC	-1- TISSUE SPECIFICITY: STAGE 10 EMBRYOS EXHIBIT INTENSE EXPRESSION		
CC	IN EPITHELIAL CELLS.		
CC	STAGE 14 EMBRYOS SHOW EXPRESSION IN THE HANDGUT		
CC	(AT THE APICAL POLDS OF CELL-CELL BOUNDARIES), AT THE APICAL		
CC	JUNCTIONS OF TRACHEAL CELLS AND IN THE DORSAL LONGITUDINAL TRUNK		
CC	IN STAGE 16 EMBRYOS THE GLIAL MIDLINE CELLS OF THE CENTRAL		
CC	NERVOUS SYSTEM SHOW STRONG EXPRESSION.		
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.		
CC	-1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.		
DR	EMBL; D28749; BAA05942.1; .		
DR	PROSITE; PS00232; CADHERIN. 5.		
DR	PROSITE; PS00022; EGF_1; 1.		
DR	PROSITE; PS01186; EGF_2; 1.		
DR	FLYBASE; FBgn0003391; shg.		
DR	PFAM; PF00028; cadherin_7.		
DR	PFAM; PF01049; Cadherin_C-term; 1.		
KW	Cell adhesion; laminin_G; 1.		
KW	Signal; zymogen.		
FT	PROPEP	?	POTENTIAL.
FT	CHAIN	261	POTENTIAL.
FT	DOMAIN	262 1507	DE-CADHERIN PROTEIN.
FT	TRANSMEM	262 1328	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1329 1349	POTENTIAL.
FT	DOMAIN	1350 1507	CYTOPLASMIC (POTENTIAL).
FT	REPEAT	302 412	EGF-LIKE.
FT	REPEAT	413 522	CADHERIN 1 (CR1).
FT	REPEAT	523 623	CADHERIN 2 (CR2).
FT	REPEAT	624 733	CADHERIN 3 (CR3).
FT	REPEAT	734 836	CADHERIN 4 (CR4).
FT	REPEAT	837 937	CADHERIN 5 (CR5).
FT	CARBOHYD	317 317	POTENTIAL.
FT	CARBOHYD	466 466	POTENTIAL.
FT	CARBOHYD	552 552	POTENTIAL.
FT	CARBOHYD	766 766	POTENTIAL.
FT	CARBOHYD	949 949	POTENTIAL.
FT	CARBOHYD	983 983	POTENTIAL.
FT	CARBOHYD	999 999	POTENTIAL.
FT	CARBOHYD	1073 1073	POTENTIAL.
FT	CARBOHYD	1145 1145	POTENTIAL.

```

FT CARBOHYD 1274 1274 POTENTIAL.
FT CARBOHYD 1290 1290 POTENTIAL.
SQ SEQUENCE 1507 AA; 169854 MW; 911CC772 CRC32;

Query Match          56.8%; Score 50; DB 5; Length 1507;
Best Local Similarity 46.7%; Pred. No. 6;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY      1 CECNIRKVDVNDNF 15
       | : ::|::||| |
DB      179 CTFTNTIEDINDNP 193

RESULT   7
ID       055134 PRELIMINARY; PRT; 1180 AA.
AC       055134;
DT       01-JUN-1998 (TREMBLrel. 06, Created)
DR       01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DE       01-MAY-1999 (TREMBLrel. 10, Last annotation update)
PE       PROTOCADHERIN-4.
OS       Mus musculus (Mouse).
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN       [1]
RP       SEQUENCE FROM N.A.
RC       TISSUE-BRAIN CAPILLARY;
RA       TELC P.;
RL       Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
CC       -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
DR       EMBL: Y08715; CAAG6965.1; -.
PE       PFAM: PF00028; cadherin_5.
DR       PROSITE: PS00232; CADHERIN; 5.
KW       Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
SQ       SEQUENCE 1180 AA; 128672 MW; 48C0883F CRC32;

Query Match          56.8%; Score 50; DB 11; Length 1180;
Best Local Similarity 66.7%; Pred. No. 4;7;
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 CECNIRKVDVNDNF 15
       | : ||| ||||| |
DB      336 CKYLIRKLVVDNDNP 350

RESULT   8
ID       013737 PRELIMINARY; PRT; 222 AA.
AC       013737;
DT       01-JAN-1998 (TREMBLrel. 06, Created)
DR       01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE       01-MAY-1989 (TREMBLrel. 10, Last annotation update)
PE       HYPOTHELTICAL 24.9 KD PROTEIN C16B8.02 IN CHROMOSOME 1.
OS       SPAC16B8.02.
OC       Schizosaccharomyces pombe (Fission yeast).
OC       Eukaryota; Fungi; Ascomycota; Archaesporium;
          Schizosaccharomycetales; Schizosaccharomycetaceae;
          Schizosaccharomyces.
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN-972;
RA       OLIVER K., HARRIS D., BARRELL B.G., RAJANDREAM M.A., WOOD V.;
RL       Submitted (AUG-1997) to the EMBL/Genbank/DBJ databases.
CC       -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
          -1- SIMILARITY: TO YEAST YGL010W AND SOME, TO N.CRASSA ATP-6.
DR       EMBL: Z98529; CAB1031.1; -.
KW       Hypothetical protein; Transmembrane.
RN       TRANSMEM 27-- 47
FT       TRANSMEM 65 85 POTENTIAL.
FT       TRANSMEM 105 125 POTENTIAL.
FT       TRANSMEM 135 135 POTENTIAL.
```





RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSET T., COOPER J., COULSON A.,  
 RA CRATON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RITKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STULTON J.,  
 RA THIRRY-MING J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RL Nature 368:32-38(1994).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: 283119; CAB05582.1; -.  
 DR PFAM: PF00028; cadherin; 3.  
 DR PROSITE: PS00232; CADHERIN; 1.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 915 AA; 101482 MW; A25F734 CRC32;

QY Query Match 55.7%; Score 49; DB 5; Length 915;  
 Db Best Local Similarity 81.8%; Pred. No. 5.3; Indels 0; Gaps 0;  
 Matches 9; Conservative 1; Mismatches 1;

QY 5 IKYKVDNDNFP 15  
 Db 625 VKKYVDNDNFP 635  
 RESULT 13  
 ID 075287 PRELIMINARY; PRT; 824 AA.  
 AC 075287;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE KIA0345-LIKE 12.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA KIMMERLY W., BONDOC M., CHENG J., CONNOLLY K.S., GUNNING K.M.,  
 RA KADNER K., MIGUEL T., MILLER C., PITLICK S., POLLARD M., ROJESKI H.,  
 RA SUBRAMANIAN S., MARTIN C.H.;  
 RT "Sequencing of human chromosome 5";  
 RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA RICKE D.O.;  
 RT "Large Scale Sequence Analysis and Annotation with the Sequence  
 RT Comparison Analysis (SCAN) System";  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AC005609; AAC34324.1; -.  
 DR PFAM: PF00028; cadherin; 5.  
 DR PROSITE: PS00232; CADHERIN; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 824 AA; 88866 MW; 56C50E1B CRC32;

QY Query Match 54.5%; Score 48; DB 4; Length 824;  
 Db Best Local Similarity 46.7%; Pred. No. 6.9; Indels 0; Gaps 0;  
 Matches 7; Conservative 5; Mismatches 3;

RESULT 14  
 ID 035542 PRELIMINARY; PRT; 784 AA.  
 AC 035542;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE CADHERIN-5  
 OS Mus musculus (Mouse)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 97364256.  
 RA MAISTOSH N., TODA K., HORIGUCHI Y., TANAKA T., NAKAGAWA S.,  
 RA TAKEICHI M., IMAMURA S.;  
 RT "In vivo evidence of the critical role of cadherin-5 in murine  
 RT vascular integrity.";  
 RL Proc. Assoc. Am. Physiologists 109:362-371(1997).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: D63942; BAA22617.1; -.  
 DR PFAM: PF00028; cadherin; 5.  
 DR PFAM: PF01049; Cadherin\_C term; 1.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 784 AA; 87902 MW; 68D4FB78 CRC32;

QY Query Match 54.5%; Score 48; DB 11; Length 784;  
 Db Best Local Similarity 63.6%; Pred. No. 6.6; Indels 0; Gaps 0;  
 Matches 7; Conservative 4; Mismatches 0;

QY 5 IKYKVDNDNFP 15  
 Db 244 IRLIEDNDNFP 254  
 RESULT 15  
 ID 086689 PRELIMINARY; PRT; 947 AA.  
 AC 086689;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE CNR1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE: 98318234.  
 RA KOHMOURA N., SENZAKI K., HAMADA S., KAI N., YASUDA R., WATANABE M.,  
 RA ISHII H., YASUDA M., MISHINA M., YAGI T.;  
 RT "Diversity revealed by a novel family of cadherins expressed in  
 RT neurons at a synaptic complex.";  
 RL Neuron 20:1137-1151(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: D86916; BAA29045.1; -.  
 DR PFAM: PF00028; cadherin; 6.  
 DR PROSITE: PS00232; CADHERIN; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 947 AA; 103143 MW; AB58B52C CRC32;

QY Query Match 54.5%; Score 48; DB 11; Length 947;  
 Db Best Local Similarity 46.7%; Pred. No. 7.9; Indels 0; Gaps 0;  
 Matches 7; Conservative 5; Mismatches 3;

Sat Jan 15 11:45:04 2000

us-08-991-628-5.rspt

Search completed: January 12, 2000, 23:15:23  
Job time: 186 sec



A:Residues: 1-1049 <RUX>  
A:Cross-references: EMBL:X56654; NID:930505; PIDN:CAA39976.1; PID:930506  
R:Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Atlioclu, P.; Poynter, D.; Arneemann, J.; R  
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991  
A:Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions.  
A:Reference number: A39706; MUID:91271279  
A:Accession: A39706  
A:Molecule type: mRNA  
A:Residues: 24-1049 <WHE>  
A:Cross-references: GB:X56654  
R:Nilles, L.A.; Perry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.  
J. Cell Sci. 99, 809-821, 1991  
A:Title: Structural analysis and expression of human desmoglein: a cadherin-like compone  
A:Reference number: A61254; MUID:92121251  
A:Accession: A61254  
A:Molecule type: mRNA  
A:Residues: 26-1049 <NLT>  
R:Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; Kling, I.A.; Magee  
Biochem. Soc. Trans. 19, 1060-1064, 1991  
A:Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily  
A:Reference number: A61279; MUID:92175187  
A:Accession: A61279  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-55 <WH3>  
C:Genetics:  
A:Gene: GDB:DSG1  
A:Cross-references: GDB:126563; OMIM:125670  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-1049/Product: desmoglein #status predicted <MNT>  
F:50-548/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-269/Domain: cadherin repeat homology <CR2>  
F:272-385/Domain: cadherin repeat homology <CR3>  
F:392-493/Domain: cadherin repeat homology <CR4>  
F:509-530/Region: serine/threonine-rich  
F:549-569/Domain: transmembrane #status predicted <TM>  
F:572-1049/Domain: intracellular #status predicted <INT>  
F:840-869/Domain: desmoglein repeat <DG1>  
F:870-899/Domain: desmoglein repeat <DG2>  
F:900-927/Domain: desmoglein repeat <DG3>  
F:928-956/Domain: desmoglein repeat <DG4>  
F:969-1019/Region: glycine/serine-rich  
F:110/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.5%; Score 64; DB 1; Length 1049;  
Best Local Similarity 80.0%; Pred. No. 0.0014;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATOKITRISGVGID 15  
Db 79 ANQVTVTRISGVGID 93

RESULT 3  
10BOG1  
desmoglein 1 precursor - bovine  
N:Alternate names: desmoglein BDGM  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: S14603; A38672; A37871; A48173; S24412  
R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.  
submitted to the EMBL Data Library, March 1991  
A:Description: Complete sequence of the desmoglein precursor and evidence for the existe  
A:Reference number: S14603  
A:Accession: S14603  
A:Molecule type: mRNA  
A:Residues: 1-1043 <KOC>

A:Cross-references: EMBL:X58466; NID:9306; PIDN:CAA41380.1; PID:9307  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Eur. J. Cell Biol. 55, 200-208, 1991  
A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypept  
A:Reference number: A38872; MUID:92037656  
A:Accession: A38872  
A:Molecule type: mRNA  
A:Residues: 1-87,968-1043 <KOC>  
A:Cross-references: GB:S64268; GB:S64270  
R:Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Marabe, M.; Cowlin, P.  
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990  
A:Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion  
A:Reference number: A37785; MUID:91097553  
A:Accession: A37785  
A:Molecule type: mRNA  
A:Residues: 44-123, 'V', 125-493 <GOC>  
A:Cross-references: GB:M58165; NID:9162966; PIDN:AAA62709.1; PID:9552318  
R:Zimbelmann, R.  
submitted to the EMBL Data Library, February 1991  
A:Reference number: S38721  
A:Accession: S38721  
A:Molecule type: mRNA  
A:Residues: 44-1043 <ZIM>  
A:Cross-references: EMBL:X57784; NID:9436061; PIDN:CAA40930.1; PID:9436062  
R:Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.  
Eur. J. Cell Biol. 53, 1-12, 1990  
A:Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m  
A:Reference number: A48173; MUID:91168965  
A:Accession: A48173  
A:Molecule type: mRNA  
A:Residues: 44-1001, 'ADPPSAT' <KOC>  
A:Cross-references: GB:X57784  
A:Note: this sequence has been revised in references A38872 and S38721  
C:Genetics:  
A:Gene: DSG1  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-1043/Product: desmoglein #status predicted <MNT>  
F:50-548/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-269/Domain: cadherin repeat homology <CR2>  
F:272-385/Domain: cadherin repeat homology <CR3>  
F:392-491/Domain: cadherin repeat homology <CR4>  
F:549-574/Domain: transmembrane #status predicted <TM>  
F:575-1043/Domain: intracellular #status predicted <INT>  
F:846-875/Domain: desmoglein repeat <DG1>  
F:876-905/Domain: desmoglein repeat <DG2>  
F:906-933/Domain: desmoglein repeat <DG3>  
F:934-962/Domain: desmoglein repeat <DG4>  
F:963-1012/Region: glycine/serine-rich  
F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:180/96/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 86.5%; Score 64; DB 1; Length 1043;  
Best Local Similarity 80.0%; Pred. No. 0.0014;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATOKITRISGVGID 15  
Db 79 ANQVTVTRISGVGID 93

RESULT 4  
T10963  
phosphoribosylformylglycinamide cyclo-ligase (EC 6.3.3.1) - compea  
N:Alternate names: AIR synthase; AIRS; phosphoribosylaminoimidazole synthetase  
C:Species: Vigna unguiculata (cowpea)  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 24-Sep-1999  
C:Accession: T10963  
R:Smith, P.M.; Mann, A.J.; Gogglin, D.E.; Atkins, C.A.



A:Molecule type: DNA  
 A:Residues: 1-441 <ROU>  
 A:Cross-references: EMBL:AC003000; NID:g2642152; PID:g2642153  
 A:Experimental source: cultivar Columbia  
 A:Genetics:  
 A:Map position: 2  
 A:Note: T517.1

Query Match 52.7%; Score 39; DB 2; Length 441;  
 Best Local Similarity 54.5%; Pred. No. 21;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 OKITRISGVC 13  
 ||: ||: ||: ||  
 Db 368 OKLRYRVNGAC 378

RESULT 9  
 JC5830  
 myocillin - human  
 C:Species: Homo sapiens (man)  
 C>Date: 20-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 17-Mar-1999  
 C:Accession: JC5830  
 R:Kubota, R.; Kudoh, J.; Mashima, Y.; Asakawa, S.; Minoshima, S.; Hejtmancik, J.F.; Oguc  
 Blochem. Biophys. Res. Commun. 242, 396-400, 1998  
 A:Title: Genomic organization of the human myocillin gene (MYOC) responsible for primary  
 A:Reference number: JC5830; MUID:98113364  
 A:Accession: JC5830  
 A:Molecule type: DNA  
 A:Residues: 1-504 <KUB>  
 A:Cross-references: DDBJ:AB006686  
 C:Comment: This cytoskeletal protein is involved in the morphogenesis of the basal body  
 ucoma.  
 C:Genetics:  
 A:Gene: myoc  
 A:Introns: 202/1; 244/1

Query Match 52.7%; Score 39; DB 2; Length 504;  
 Best Local Similarity 57.1%; Pred. No. 24;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 TOKTIRISGVID 15  
 ||: ||: ||: ||  
 Db 281 TOETTRIDTVGTD 294

RESULT 10  
 A46016  
 thrombospondin 3 - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A46016; A44124; I55398; S68188; S72433  
 R:Bornstein, P.; Devarayalu, S.; Edelhoff, S.; Distech, C.M.  
 Genomics 15, 607-613, 1993  
 A:Title: Isolation and characterization of the mouse thrombospondin 3 (Thbs3) gene.  
 A:Reference number: A46016; MUID:93224149  
 A:Accession: A46016  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-956 <BOR>  
 A:Cross-references: GB:I04302; NID:g202200; PID:AAA0497.1; PID:g202201  
 A:Note: sequence extracted from NCBI backbone (NCBIN:129415; NCBIPI:129416)  
 R:Vos, H.L.; Devarayalu, S.; de Vries, Y.; Bornstein, P.  
 J. Biol. Chem. 267, 12192-12196, 1992  
 A:Title: Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene family.  
 A:Reference number: A44124; MUID:92291102  
 A:Accession: A44124  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 517-956 <VOS>  
 A:Cross-references: GB:M86620; NID:g201987; PID:AAA0430.1; PID:g201989

A:Experimental source: BAB/C  
 A:Note: sequence extracted from NCBI backbone (NCBIN:106634; NCBIN:106636; NCBIN:1066  
 )  
 R:Gaber, A.N.; Lin, Z.; Wolf, F.W.; O'Shea, K.S.; Lawler, J.; Dixit, V.M.  
 J. Biol. Chem. 269, 1262-1269, 1994  
 A:Title: Thrombospondin 3 is a developmentally regulated heparin binding protein.  
 A:Reference number: I55398; MUID:94117438  
 A:Accession: I55398

A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-432, 'F', 434-562, 'D', 564-719, 'E', 721-870, 'W', 872-956 <RES>  
 A:Cross-references: GB:I24434; NID:9402718; PID:AAA0433.1; PID:9402719  
 R:Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D.F.

FEBS Lett. 387, 36-41, 1996  
 A:Title: Expression and initial characterization of recombinant mouse thrombospondin  
 A:Reference number: S68787; MUID:96234006  
 A:Accession: S68788  
 A:Molecule type: protein

A:Residues: 'X', 23-28, 'X', 30, 'X', 32 <CHE>  
 R:Collins, M.; Bornstein, P.  
 Nucleic Acids Res. 24, 3661-3669, 1996  
 A:Title: SPI-binding elements, within the common metaxin-thrombospondin 3 intergenic  
 A:Reference number: S72433; MUID:97025352  
 A:Accession: S72433

A:Molecule type: DNA  
 A:Residues: 1-26 <COL>  
 A:Cross-references: EMBL:U66257; NID:g1575552; PID:RAC52819.1; PID:g1575554  
 C:Genetics:  
 A:Gene: Thbs-3

C:Complex: homopentamer, disulfide linked  
 C:Superfamily: thrombospondin 3; Egf homology  
 C:Keywords: calcium binding; glycoprotein; homopentamer  
 F:310,407,644,937/Binding site: carbohydrate, (Asn) (covalent) #status predicted

Query Match 51.4%; Score 38; DB 1; Length 956;  
 Best Local Similarity 87.5%; Pred. No. 72;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 8 RISGVGID 15  
 ||: ||: ||: ||  
 Db 352 RVSGVGID 359

RESULT 11  
 D71416  
 probable PDR5-like ABC transporter - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 A:Vareity: Columbia  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: D71416  
 R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
 P.; Wedler, H.; Wedler, E.; Wambutt, R.; Welzenegger, T.; Pohl, T.M.; Terry, N.; G  
 Navarath, T.; Hempel, S.; Kotter, P.; Ertlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
 Nature 391, 485-488, 1998  
 A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgaden  
 erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reichman, S.;  
 C.; Chalvatzki, N.  
 A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
 A:Reference number: A71400; MUID:98121113  
 A:Accession: D71416  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-1177 <BEV>  
 A:Cross-references: GB:I297338; NID:g2244870; PID:e336915; PID:g2244881  
 C:Genetics:  
 A:Map position: 4COP9-4G3845  
 C:Superfamily: Arabidopsis thaliana probable PDR5-like ABC transporter; ATP-binding c  
 F:439-653/Domain: ATP-binding cassette homology <ABC2>

Query Match 51.4%; Score 38; DB 1; Length 1177;

Best Local Similarity 69.2%; Pred. No. 90;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 OKITIRISGVGID 15  
||| ||| |||  
Db 1042 OKIRKRIIDEVGI 1054

## RESULT 12

157009  
3-Hydroxy-3-methylglutaryl coenzyme A lyase - mouse  
C:Species: Mus sp. (mouse)  
C:Date: 02-Aug-1996 #sequence\_revision 02-Aug-1996 #text\_change 22-Jun-1999  
C:Accession: 157009  
R:Wang, S.; Nadeau, J.H.; Duncan, A.; Robert, M.F.; Fontaine, G.; Schappert, K.; Johnson  
Mamm. Genome 4, 382-387, 1993  
A:Title: 3-Hydroxy-3-methylglutaryl coenzyme A lyase (HL): cloning and characterization  
A:Reference number: 157009; MWID:93364116  
A:Accession: 157009  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-325 <RES>  
A:Cross-references: GB:S65036; NID:g409498; PIDN:AAB27965.1; PID:g409499  
C:Genetics:  
A:Gene: Hmgcl  
C:Superfamily: hydroxymethylglutaryl-CoA lyase

Query Match 51.4%; Score 38; DB 2; Length 325;  
Best Local Similarity 42.9%; Pred. No. 22;

Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATOKITIRISGVGI 14  
||: ||: ||: ||:  
Db 277 ATEDLVYMLNGIGI 290

## RESULT 13

B45470  
hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4) - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 25-Mar-1998  
R:Mitchell, G.A.; Robert, M.F.; Hruz, P.W.; Wang, S.; Fontaine, G.; Behnke, C.E.; Mende-  
J. Biol. Chem. 268, 4376-4381, 1993  
A:Title: 3-Hydroxy-3-methylglutaryl coenzyme A lyase (HL): Cloning of human and chicken  
A:Reference number: A45470; MWID:93179448  
A:Accession: B45470  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-298 <MIT>  
A:Cross-references: GB:I07034  
C:Superfamily: hydroxymethylglutaryl-CoA lyase  
C:Keywords: carbon-carbon lyase; mitochondrion; oxo-acid-lyase

Query Match 51.4%; Score 38; DB 2; Length 298;  
Best Local Similarity 42.9%; Pred. No. 20;

Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AFOKITIRISGVGI 14  
||: ||: ||: ||:  
Db 250 ATEDLVYMLNGIGI 263

## RESULT 14

T02655  
hydroxymethylglutaryl-CoA lyase (EC 4.1.3.4) - Arabidopsis thaliana  
N:Alternate names: protein F12C20.16  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 22-Jun-1999  
C:Accession: T02655  
R:Rounsley, S.D.; Romling, C.M.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sy

submitted to the EMBL Data Library, August 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F12C20 genomic sequence.  
A:Reference number: Z14685

A:Accession: T02655  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-433 <ROU>  
A:Cross-references: EMBL:AC005168; NID:g3426033; PIDN:AAC32247.1; PID:g3426048  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:Intons: 55/2; 124/3; 184/3; 211/3; 266/2; 312/1; 350/3  
A:Note: F12C20.16  
C:Superfamily: hydroxymethylglutaryl-CoA lyase  
C:Keywords: carbon-carbon lyase; oxo-acid-lyase

Query Match 51.4%; Score 38; DB 2; Length 433;  
Best Local Similarity 35.7%; Pred. No. 31;  
Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATOKITIRISGVGI 14  
||: ||: ||: ||:  
Db 377 ATEDLVYMLNGIGI 390

## RESULT 15

E70367  
phosphoribosylformylglycinamide cyclo-lyase - Aquifex aeolicus  
C:Species: Aquifex aeolicus  
C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 16-Jul-1999  
C:Accession: E70367  
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
V. Nature 392, 353-358, 1998  
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A:Reference number: A70300; MWID:98196666  
A:Accession: E70367  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-331 <AOE>  
A:Cross-references: GB:AE000706; NID:g2983327; PIDN:AAC06918.1; PID:g2983329; GB:AE00  
A:Experimental source: strain VFS  
C:Genetics:  
A:Gene: purM  
C:Superfamily: phosphoribosylformylglycinamide cyclo-lyase; phosphoribosylformylgl  
F:4-317/Domains: phosphoribosylformylglycinamide cyclo-lyase homology <PFCt>

Query Match 51.4%; Score 38; DB 2; Length 331;  
Best Local Similarity 63.6%; Pred. No. 23;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 ITYRISGVGID 15  
||| ||| |||  
Db 2 VTYRSAGVID 12

Search completed: January 7, 2000, 08:52:37  
Job time: 2267 sec





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:19:08 ; Search time 28.55 Seconds  
(without alignments)  
15.282 Million cell updates/sec

Title: US-08-991-628-1  
Perfect score: 74  
Sequence: 1 ATOKITRISGVGID 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38:\*

Word size: 0

Number of hits that pass the threshold : 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74	100.0	999	1	DSG3_HUMAN
2	64	86.5	1043	1	DSG1_BOVIN
3	64	86.5	1049	1	DSG1_HUMAN
4	44	59.5	388	1	PUR5_VICUN
5	42	56.8	897	1	BCAL_GLOBE
6	42	56.8	350	1	CBS_HUMAN
7	39	52.7	504	1	MYOC_HUMAN
8	38	51.4	140	1	HMGL_BOVIN
9	38	51.4	298	1	HMGL_CHICK
10	38	51.4	325	1	HMGL_MOUSE
11	38	51.4	325	1	HMGL_RAT
12	37	50.0	956	1	TSP3_MOUSE
13	37	50.0	560	1	CBS_RAT
14	37	50.0	1117	1	DSG2_HUMAN
15	37	50.0	325	1	HMGL_HUMAN
16	37	50.0	1363	1	VGL2_CVBF
17	37	50.0	1363	1	VGL2_CVBF
18	37	50.0	1363	1	VGL2_CVBF
19	37	50.0	1363	1	VGL2_CVBF
20	37	50.0	1363	1	VGL2_CVBF
21	37	50.0	1363	1	VGL2_CVBF
22	37	50.0	1121	1	VGL2_CVBF
23	37	50.0	1218	1	YK83_YEAST
24	36	48.6	890	1	BCN5_CLOPE
25	36	48.6	712	1	CADD_CHICK
26	36	48.6	761	1	METE_BACAU
27	36	48.6	690	1	NCPR_PHAU
28	36	48.6	247	1	OPAG_NEICO
29	36	48.6	740	1	PFL_CLOPA
30	36	48.6	335	1	PUR5_ARATH
31	36	48.6	323	1	YF78_HAEIN
32	35	47.3	420	1	ALAU_HUMAN
33	35	47.3	859	1	ABR_HUMAN
34	35	47.3	985	1	AGLU_ASPNG
35	35	47.3	127	1	CS60_CHOCR
36	35	47.3	882	1	CAD1_MOUSE
37	35	47.3	884	1	CAD1_MOUSE
38	35	47.3	913	1	CAD4_CHICK
39	35	47.3	916	1	CAD4_HUMAN

40	35	47.3	913	1	CAD4_MOUSE
41	35	47.3	509	1	CRU3_BRANA
42	35	47.3	893	1	DSG1_BOVIN
43	35	47.3	894	1	DSG1_HUMAN
44	35	47.3	863	1	DSG2_BOVIN
45	35	47.3	607	1	G6PI_TRYBB
					P13377 trypanosoma

## ALIGNMENTS

RESULT	ID	STANDARD	PRT	999 AA
1	DSG3_HUMAN			
1	DSG3_HUMAN			
AC	P32926			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-OCT-1993 (Rel. 27, Last sequence update)			
DT	01-NOV-1997 (Rel. 35, Last annotation update)			
DE	DESMOGLIN 3 PRECURSOR (130 KD PEMPHIGUS VULGARIS ANTIGEN) (PVA).			
GN	DSG3.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;			
CC	Eutheria; Primates; Catarrhini; Homiidae; Homo.			
RP	SEQUENCE FROM N.A.			
FX	MEDLINE; 92069753.			
RA	AMAGAI M., KLAUS-KOVTON V., STANLEY J.R.,			
RT	"Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a disease of cell adhesion."			
RL	Cell 67:869-877(1991).			
CC	-1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS. INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENTS MEDIATING CELL-CELL ADHESION.			
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ESOPHAGUS AND CARCINOMAS.			
CC	-1- DOMAIN: CALCIDIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).			
CC	-1- DISEASE: PEMPHIGUS VULGARIS (PV) IS A POTENTIALLY LETHAL SKIN DISEASE IN WHICH EPIDERMAL BLISTERS OCCUR AS THE RESULT OF THE LOSS OF CELL-CELL ADHESION CAUSED BY THE ACTION OF AUTOANTIBODIES AGAINST DSG3.			
CC	-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.			
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION - THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL ENTITIES REQUIRES A LICENSE AGREEMENT (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; M76482; AAA60230.1; -			
DR	PIR; A41088; IYH03.			
DR	HSSP; P09803; 1EDH.			
DR	MIT; 169615; -			
DR	PFAM; PF00028; cadherin; 4.			
DR	PROSITE; PS00233; CADHERIN; 3.			
KW	Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;			
KW	Calcium-binding; Repeat.			
FT	STIGNAL 1			
FT	PROPEP 24			
FT	CHAIN 30			
FT	DOMAIN 50			
FT	DOMAIN 616			
FT	TRANSMEM 640			
FT	DOMAIN 641			
FT	REPEAT 50			
FT	REPEAT 159			
FT	REPEAT 269			
FT	REPEAT 383			
FT	REPEAT 386			
FT	REPEAT 499			
FT	REPEAT 910			
FT	REPEAT 936			
FT	REPEAT 966			
FT	CARBOHYD 110			
	POTENTIAL.			
	DESMOGLIN 3.			
	EXTRACELLULAR (POTENTIAL).			
	POTENTIAL.			
	CYTOPLASMIC (POTENTIAL).			
	CADHERIN 1.			
	CADHERIN 2.			
	CADHERIN 3.			
	CADHERIN 4.			
	DESMOGLIN REPEAT 1.			
	DESMOGLIN REPEAT 2.			
	POTENTIAL.			

FT CARBOHYD 180 180 POTENTIAL.  
 FT CARBOHYD 459 459 POTENTIAL.  
 FT CARBOHYD 545 545 POTENTIAL.  
 SQ SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;

Query Match 100.0%; Score 74; DB 1; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATOKITRISGVGID 15  
 Db 79 ANQOVTYRISGVGID 93

RESULT 2  
 ID DSG1\_BOVIN STANDARD; PRT; 1043 AA.  
 AC 003763;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (Dg1).  
 GN DSG1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=MUZZLE EPITHELIUM;  
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;  
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RN SEQUENCE OF 44-1043 FROM N.A.  
 RC TISSUE=MUZZLE EPITHELIUM;  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
 RX MEDLINE: 91168965.  
 RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,  
 RA ZIMBELMANN R., FRANK W.W.;  
 RT "Identification of desmoglein, a constitutive desmosomal  
 RT glycoprotein, as a member of the cadherin family of cell adhesion  
 RT molecules.";  
 RL Eur. J. Cell Biol. 53:1-12(1990).  
 RN [3]  
 RN REVISIONS, AND SEQUENCE OF 101-123.  
 RX MEDLINE: 92037656.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
 RA FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT polypeptide and identification of a second type of desmoglein gene.";  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 RN [4]  
 RN SEQUENCE OF 44-493 FROM N.A.  
 RX MEDLINE: 91097553.  
 RA GOODWIN L., HILL J.E., RAYNOR K., RASZI L., MANABE M., COWIN P.;  
 RT "Desmoglein shows extensive homology to the cadherin family of cell  
 RT adhesion molecules.";  
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).  
 RN [5]  
 RN FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.  
 CC DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>)

CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).

CC EMBL: X58466; CAA41380.1;  
 CC EMBL: X57784; CAA40930.1;  
 CC EMBL: M58165; AAB62709.1;  
 CC PIR: S14603; IJBOG1.  
 CC HSSP: P09803; IEDH.  
 CC PIRAT: PF00026; cadherin. 3.  
 CC PROSITE: PS00332; CADHERIN. 2.  
 CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 CC Calcium-binding; Repeat.  
 FT SIGNAL 1 23  
 FT PROPEP 24 49  
 FT CHAIN 50 1043  
 FT DOMAIN 50 548  
 FT TRANSMEM 549 573  
 FT DOMAIN 574 1043  
 FT REPEAT 50 158  
 FT REPEAT 159 270  
 FT REPEAT 271 385  
 FT REPEAT 386 498  
 FT REPEAT 819 845  
 FT REPEAT 846 875  
 FT REPEAT 876 905  
 FT REPEAT 906 933  
 FT REPEAT 934 962  
 FT DOMAIN 963 1012  
 FT CARBOHYD 110 110  
 FT CARBOHYD 180 180  
 FT CARBOHYD 496 496  
 FT CONFLICT 124 124  
 SQ SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;  
 POTENTIAL.  
 POTENTIAL.  
 I->V (IN REF. 4).

Query Match 86.5%; Score 64; DB 1; Length 1043;  
 Best Local Similarity 80.0%; Pred. No. 0.00098;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATOKITRISGVGID 15  
 Db 79 ANQOVTYRISGVGID 93

RESULT 3  
 ID DSG1\_HUMAN STANDARD; PRT; 1049 AA.  
 AC 002413;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (Dg1).  
 GN DSG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=KERATINOCYTES;  
 RX MEDLINE: 91271279.  
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALLOTIS P., POYNTER D.,  
 RA ARNDMAN J., RUTMAN A.J., PIDSELY S.C., WATT F.M., REES D.A.,  
 RA BUXTON R.S., MAGEE A.I.;  
 RT "Desmosomal glycoprotein Dg1, a component of intercellular desmosome  
 RT junctions, is related to the cadherin family of cell adhesion  
 RT molecules.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).  
 RN [2]  
 RN FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.  
 CC DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).

```

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X56654; CA39976.1; -.
CC PIR: S16906; IJHUG1.
CC HSSP: P09803; LEDH.
CC MIM: 125670; -.
CC PFM: PFO0028; cadherin. 4.
CC PROSITE: PS00232; CADHERIN. 2.
CC K1M Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;
CC Calcium-binding; Repeat.
CC FT SIGNAL 1 23 POTENTIAL.
CC FT PROPEP 24 49 POTENTIAL.
CC FT CHAIN 30 1049 DESMOGLEIN 1.
CC FT DOMAIN 50 545 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 546 570 POTENTIAL.
CC FT DOMAIN 571 1049 CYTOPLASMIC (POTENTIAL).
CC FT REPEAT 50 158 CADHERIN 1.
CC FT REPEAT 159 270 CADHERIN 2.
CC FT REPEAT 271 385 CADHERIN 3.
CC FT REPEAT 386 497 CADHERIN 4.
CC FT REPEAT 497 839 DESMOGLEIN REPEAT 1.
CC FT REPEAT 840 869 DESMOGLEIN REPEAT 2.
CC FT REPEAT 870 899 DESMOGLEIN REPEAT 3.
CC FT REPEAT 900 927 DESMOGLEIN REPEAT 4.
CC FT REPEAT 928 936 DESMOGLEIN REPEAT 5.
CC FT DOMAIN 969 1019 GLY/SER-RICH.
CC FT CARBOHYD 36 36 POTENTIAL.
CC FT CARBOHYD 110 110 POTENTIAL.
CC FT CARBOHYD 180 180 POTENTIAL.
CC SQ SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;

Query Match 86.5%; Score 64; DB 1; Length 1049;
Best Local Similarity 80.0%; Pred. No. 0.00099;
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATOKITRISGVGID 15
Db 79 ANOCVTRISGVGID 93

RESULT 4
PUR5_VIGUN STANDARD; PRT; 388 AA.
AC P52424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOSPHORIBOSYL-AMINOIMIDAZOLE CYCLO-LIGASE PRACUSOR (EC 6.3.3.1)
DE (AIRS) (PHOSPHORIBOSYL-AMINOIMIDAZOLE SYNTHETASE) (AIR SYNTHASE).
GN PNR5.
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; euroids I; Fabales; Fabaceae; Papilionoideae;
OC Vigna.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-CV. VITA 3; TISSUE-ROOT NODULES;
CC SMITH P.M.C., MANN A.J., HALL D.J., ATKINS C.A.;
CC Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 5'-PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE -
CC ADP + ORTHOPHOSPHATE + 5'-PHOSPHORIBOSYL-5-AMINOIMIDAZOLE.
CC -1- PATHWAY: FIFTH STEP IN DE NOVO PURINE BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST.

```

```

CC -1- SIMILARITY: TO OTHER AIRS FROM BACTERIA AND EUKARYOTES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U30895; AAC14578.1; -.
CC PFM: PFO0586; AIRS; 1.
CC K1M Purine biosynthesis; Ligase; Chloroplast; Transist peptide.
CC FT TRANST 1 388 CHLOROPLAST (POTENTIAL).
CC FT CHAIN 1 388 PHOSPHORIBOSYL-FORMYLGLYCINAMIDINE CYCLO-
CC LIGASE.
CC SQ SEQUENCE 388 AA; 40427 MW; 895D3159 CRC32;

Query Match 59.5%; Score 44; DB 1; Length 388;
Best Local Similarity 60.0%; Pred. No. 1.4;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATOKITRISGVGID 15
Db 62 ASOGITRDAGVID 76

RESULT 5
BGAL_CLOAB STANDARD; PRT; 897 AA.
AC P24131;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE).
GN CGA.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN-NCIB 2951;
CC RX MEDLINE: 91216979.
CC RA HANCOCK K.R., ROCKMAN E., YOUNG C.A., PEARCE L., MADDOX I.S.,
CC SCOTT D.B.;
CC RT "Expression and nucleotide sequence of the Clostridium acetobutylicum
CC beta-galactosidase gene cloned in Escherichia coli."
CC J. Bacteriol. 173:3084-3095(1991).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
CC -1- INDUCTION: LATE IN THE ABE (ACETONE, BUTANOL, AND ETHANOL)
CC FERMENTATION AND SUBJECT TO GLUCOSE REPRESSION.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35107; AAA23216.1; -.
CC PIR: A39405; A39405.
CC HSSP: P00722; IBGL.
CC PFM: PFO0703; Glyco_hydro_2; 1.
CC PROSITE: PS00719; GLYCOSYL_HYDROL_F2_1; 1.
CC PROSITE: PS00608; GLYCOSYL_HYDROL_F2_2; 1.
CC K1M Hydrolase; Glyco4dase.
CC FT ACT_SITE 459 459 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 503 503 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 897 AA; 105020 MW; 5CFE268B CRC32;

```

Query Match 56.8%; Score 42; DB 1; Length 897;  
 Best Local Similarity 72.7%; Ped. No. 7.5;  
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

4 KITRISGVI 14  
 |||||  
 828 KITRISGVI 838

RESULT 6  
 CBS\_HUMAN STANDARD; PRT; 550 AA.

AC P35520; O99425; (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE CYSTATHIONINE BETA-SYNTHASE (EC 4.2.1.22) (SERINE SULFHYDRASE)  
 DE (BETA-THIONASE).  
 GN CBS.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA KRAUS J.P., LE K., SWARCOOP M., OHURA T., TAHARA T., ROSENBERG L.E.,  
 RA ROOPER M.D., KOZICH V.;  
 RT "Human cystathionine beta-synthase cDNA: sequence, alternative  
 RT splicing and expression in cultured cells.";  
 RL Hum. Mol. Genet. 2:1633-1638(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LIVER;  
 RA CHASSE J.-F., PALY E., PARIS D., PAUL V., SINET P.M., KAMOUN P.,  
 RA LONDON J.;  
 RT "Genomic organization of the human cystathionine beta-synthase gene:  
 RT evidence for various CDNA's.";  
 RL Biochem. Biophys. Res. Commun. 211:826-832(1995).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC KRUGER W.D., COX D.R.;  
 RA "A yeast system for expression of human cystathionine beta-synthase:"  
 RT structural and functional conservation of the human and yeast genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:6614-6618(1994).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RA CHASSE J.-F., PAUL V., ESCANEZ R., KAMOUN P., LONDON J.;  
 RT "Human cystathionine beta-synthase: gene organization and expression  
 RT of different 5' alternative splicing.";  
 RL Mamm. Genome 8:917-921(1997).  
 [5]  
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.  
 RA KRAUS J.P., OLIVERUSOVA J., SOKOLOVA J., KRAUS E., VLCEK C.,  
 RA DE FRANCHIS R., MACLEAN K.N., BAO L., BUKOVSKA G., PATTERSON D.,  
 RA PACS V., ANSGORGE W., KOZICH V.;  
 RT "The human cystathionine beta-synthase (CBS) gene: complete sequence,  
 RT alternative splicing, and polymorphisms.";  
 RL Genomics 52:312-324(1998).  
 [6]  
 RP CHARACTERIZATION.  
 RA KRAUS J.P., PACKMAN S., FOWLER B., ROSENBERG L.E.;  
 RT "Purification and properties of cystathionine beta-synthase from  
 RT human liver. Evidence for identical subunits.";  
 RL J. Biol. Chem. 253:6523-6528(1978).  
 [7]

RP VARIANT THR-277.  
 RA MEDLINE; 93244841.  
 RA KOZICH V., KRAUS J.P.;  
 RT "Screening for mutations by expressing patient cDNA segments in E.  
 RT coli: homocystinuria due to cystathionine beta-synthase deficiency.";  
 RL Hum. Mutat. 1:113-123(1992).  
 [8]  
 RP VARIANT VAL-113 AND LEU-144.  
 RA MEDLINE; 93357760.  
 RA KOZICH V., DE FRANCHIS R., KRAUS J.P.;  
 RT "Molecular defect in a patient with pyridoxine-responsive  
 RT homocystinuria.";  
 RL Hum. Mol. Genet. 2:815-816(1993).  
 [9]  
 RP VARIANT THR-277 AND SER-306.  
 RA MEDLINE; 94108438.  
 RA HU F.L., GU Z., KOZICH V., KRAUS J.P., RAMESH V., SHIH V.E.;  
 RT "Molecular basis of cystathionine beta-synthase deficiency in  
 RT pyridoxine responsive and nonresponsive homocystinuria.";  
 RL Hum. Mol. Genet. 2:1857-1860(1993).  
 [10]  
 RP VARIANT ARG-77 AND ASN-101.  
 RA MEDLINE; 95072574.  
 RA DE FRANCHIS R., KOZICH V., MCINNIS R., KRAUS J.P.;  
 RT "Identical genotypes in siblings with different homocystinuric  
 RT phenotypes: identification of three mutations in cystathionine beta-  
 RT synthase using an improved bacterial expression system.";  
 RL Hum. Mol. Genet. 3:1103-1108(1994).  
 [11]  
 RP VARIANT GLN-124 AND ASP-130.  
 RA MEDLINE; 95152517.  
 RA MARBLE M., GRAGHTY M.T., DE FRANCHIS R., KRAUS J.P., VALLE D.;  
 RT "Characterization of a cystathionine beta-synthase allele with three  
 RT mutations in cis in a patient with B6 nonresponsive homocystinuria.";  
 RL Hum. Mol. Genet. 3:1883-1886(1994).  
 [12]  
 RP VARIANT.  
 RA MEDLINE; 95056993.  
 RA KRAUS J.P.;  
 RT "Komtewer Lecture. Molecular basis of phenotype expression in  
 RT homocystinuria.";  
 RL J. Inher. Metab. Dis. 17:383-390(1994).  
 [13]  
 RP VARIANT SER-87; GLN-124 AND MET-256.  
 RA MEDLINE; 95282779.  
 RA SEBASTIO G., SPERANDEO M.P., PANICO M., DE FRANCHIS R., KRAUS J.P.,  
 RA ANDRIA G.;  
 RT "The molecular basis of homocystinuria due to cystathionine beta-  
 RT synthase deficiency in Italian families, and report of four novel  
 RT mutations.";  
 RL Am. J. Hum. Genet. 56:1324-1333(1995).  
 [14]  
 RP VARIANT TYR-164 AND MET-370.  
 RA MEDLINE; 95362263.  
 RA KLITTMANS L.A.J., BLOW H.J., BOERS G.H.J., VAN OOST B.A.,  
 RA TRUBELS F.J.M., VAN DEN HEUVEL L.P.W.J.;  
 RT "Two novel missense mutations in the cystathionine beta-synthase gene  
 RT in homocystinuric patients.";  
 RL Hum. Genet. 96:249-250(1995).  
 [15]  
 RP VARIANT MET-167; HIS-223; THR-277; GLY-306; VAL-330 AND GLU-453.  
 RA MEDLINE; 96133289.  
 RA KRUGER W.D., COX D.R.;  
 RT "A yeast assay for functional detection of mutations in the human  
 RT cystathionine beta-synthase gene.";  
 RL Hum. Mol. Genet. 4:1155-1161(1995).  
 [16]  
 RP VARIANT LEU-289.  
 RA MEDLINE; 96057367.  
 RA SPERANDEO M.P., PANICO M., PEPE A., CANDITO M., DE FRANCHIS R.,  
 RA KRAUS J.P., ANDRIA G., SEBASTIO G.;  
 RT "Molecular analysis of patients affected by homocystinuria due to  
 RT cystathionine beta-synthase deficiency: report of a new mutation in

RT exon 8 and a deletion in intron 11." ;  
 RL J. Inherit. Metab. Dis. 18:211-214(1995).  
 RN [17]  
 RP VARIANTS GLU-363 AND SER-538.  
 RX MEDLINE; 97144192.  
 RA ARAL B., COUDE M., LONDON J., AUPELLE J., CHASSE J.-F., ZABOT M.-T.,  
 RA CHADEAUX-VEKEMANS B., KAMOUN P.;  
 RT "Two novel mutations (K384E and L539S) in the C-terminal moiety of the  
 RT cystathionine beta-synthase protein in two French pyridoxine-  
 RL responsive homocystinuria patients." ;  
 RL Hum. Mutat. 9:81-82(1997).  
 RN [18]  
 RP VARIANT ASN-443.  
 RX MEDLINE; 96331231.  
 RA KULITMAN L.A.J., BOERS G.H.J., STEVENS E.M.B., RENTER M.O.,  
 RA KRAUS J.P., TRIBELS F.J.M., VAN DEN HEUVEL L.P.W.J., BLOM H.J.;  
 RT "Defective cystathionine beta-synthase regulation by S-  
 RT adenosylmethionine in a partially pyridoxine responsive homocystinuria  
 RT patient." ;  
 RL J. Clin. Invest. 98:285-289(1996).  
 RN [19]  
 RP VARIANTS LYS-143 AND TYR-164.  
 RA GORDON R.B., COX A.J., DAWSON P.A., EMERSON B.T., KRAUS J.P.,  
 RA DUDMAN N.P.B.;  
 RT "Mutation analysis of the cystathionine beta-synthase gene: a  
 RT splicing mutation, two missense mutations and an insertion in patients  
 RT with homocystinuria." ;  
 RL Hum. Mutat. 11:332-332(1998).  
 CC -1- CATALYTIC ACTIVITY: L-SERINE + L-HOMOCYSTEINE - CYSTATHIONINE +  
 CC H(2)O.  
 CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.  
 CC -1- PATHWAY: FIRST STEP IN HOMOCYSTEINE TRANSULFURATION.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ISOFORMS; A MAJOR (SHOWN HERE) AND A  
 CC MINOR ISOFORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: IN THE ADULT STRONGLY EXPRESSED IN LIVER AND  
 CC PANCREAS. SOME EXPRESSION IN HEART AND BRAIN. WEAK EXPRESSION IN  
 CC LUNG AND KIDNEY. IN THE FETUS, EXPRESSED IN BRAIN, LIVER AND  
 CC KIDNEY.  
 CC -1- DISEASE: DEFECTS IN CBS ARE A MAJOR CAUSE OF HOMOCYSTEINURIA. THE  
 CC MOST FREQUENT SYMPTOMS INCLUDE DISLOCATED OPTIC LENSES, VASCULAR  
 CC DISORDERS (ARTERIOSCLEROSIS AND THROMBOSIS), SKELETAL  
 CC ABNORMALITIES AND MENTAL RETARDATION. PATIENTS WITH HOMOCYSTEINURIA  
 CC HAVE ELEVATED LEVELS OF HOMOCYSTEINE AND METHIONINE IN THEIR BODY  
 CC FLUID.  
 CC -1- SIMILARITY: BELONGS TO THE CYSTEINE SYNTHASE/CYSTATHIONINE BETA-  
 CC SYNTHASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 1 CBS DOMAIN.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; L19501; AAA19874.1; -;  
 DR EMBL; X82166; CAA57656.1; -;  
 DR EMBL; L14577; AAA98524.1; -;  
 DR EMBL; X88562; CAA61252.1; -;  
 DR EMBL; X91910; CAA61252.1; JOINED.  
 DR EMBL; X98811; CAA61252.1; JOINED.  
 DR EMBL; X98812; CAA61252.1; JOINED.  
 DR EMBL; X98813; CAA61252.1; JOINED.  
 DR EMBL; X98814; CAA61252.1; JOINED.

Db 295 TEORTVEEGIGID 308  
 RESULT 7  
 ID MYOC\_HUMAN STANDARD; PRT; 504 AA.  
 AC 099972; 000620;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE MYOCILIN PRECURSOR (TRABECULAR MESHWORK-INDUCED GLUCOCORTICOID  
 DE RESPONSE PROTEIN).  
 GN MYOC OR TIGR.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A. AND VARIANTS JOAG VAL-364 AND HIS-437.  
 RX MEDLINE; 97158493.  
 RA STONE E.M., FINGER J.H., ALWARD W.L.M., NGUYEN T.D., POLANSKY J.R.,  
 RA SUDEN S.L.F., NISHIMURA D., CLARK A.F., NYSTUEN A., NICHOLS B.E.,  
 RA MACKAY D.A., RITCH R., KALENAK J.W., CRAVEN E.R., SHEFFIELD V.C.;  
 RT "Identification of a gene that causes primary open angle glaucoma." ;  
 RL Science 275:668-670(1997).  
 RN [2]  
 RP REVISIONS, AND SEQUENCE OF 1-6 AND 33-37.  
 RX MEDLINE; 98165818.  
 RA NGUYEN T.D., CHEN P., HUANG W.D., CHEN H., JOHNSON D., POLANSKY J.R.;  
 RT "Gene structure and properties of TIGR, an olfactomedin-related  
 RT glycoprotein cloned from glucocorticoid-induced trabecular meshwork  
 RT cells." ;  
 RL J. Biol. Chem. 273:6341-6350(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97424389.  
 RA ORTEGO J., ESCRIBANO J., COCA-PRADOS M.;  
 RT "Cloning and characterization of substracted cDNAs from a human  
 RT ciliary body library encoding TIGR, a protein involved in juvenile  
 RT open angle glaucoma with homology to myosin and olfactomedin." ;  
 RL FEBS Lett. 413:349-353(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 97312692.  
 RA KUBOTA R., NODA S., WANG Y., MINOSHIMA S., ASAKAWA S., KUDOH J.,  
 RA MASHIMA Y., OGUCHI Y., SHIMIZU N.;  
 RT "A novel myosin-like protein (myocilin) expressed in the connecting  
 RT cillum of the photoreceptor: molecular cloning, tissue expression,  
 RT and chromosomal mapping." ;  
 RL Genomics 41:360-369(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-LEUCOCYTE;  
 RA GARCHON H.-J.;  
 RT Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA DEADMAN R.;  
 RT Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 98217378.  
 RA FINGER J.H., YING L., SWIDERSKI R.E., NYSTUEN A.M., ARBOUR N.C.,  
 RA ALWARD W.L.M., SHEFFIELD V.C., STONE E.M.;  
 RT "Characterization and comparison of the human and mouse GLCIA glaucoma  
 RT genes." ;  
 RL Genome Res. 8:377-384(1998).  
 RN [8]  
 RP VARIANTS JOAG ARG-367 AND LEU-370.  
 RX MEDLINE; 98008006.  
 RA SUZUKI Y., SHIRATO S., TANIGUCHI F., OHARA K., NISHIMAKI K., OHTA S.;  
 RT "Mutations in the TIGR gene in familial primary open-angle glaucoma

Query Match 56.88; Score 42; DB 1; Length 550;  
 Best Local Similarity 42.98; Pred. NO. 4.6;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 TOKITVRISGVIGID 15

RT in Japan." Am. J. Hum. Genet. 61:1202-1204(1997).

RN [9] VARIANTS POAG ARG-246; LEU-370; SER-477; LYS-480 AND PHE-499.

RX MEDLINE: 97472461.

RA ADAM M.F., BELMOUNEN A., BINISTI P., BREZIN A.P., VALTOT F.,

RA BECHTOLE A., DASCOTE J.-C., COPIN B., GOMEZ L., CHAVETRE A.,

RA BACH J.-F., GARCHON H.-J.;

RT "Recurrent mutations in a single exon encoding the evolutionarily

RT conserved olfactomedin-homology domain of TIGR in familial open-angle

RT glaucoma." Hum. Mol. Genet. 6:2091-2097(1997).

RN [10] VARIANTS JOAG ARG-337.

RX MEDLINE: 98027214.

RA STOIOVA D., CHILD A., BRICE G., CRICK R.P., FLECK B.W., SARPARAZI M.,

RT "Identification of a new 'TIGR' mutation in a family with juvenile-

RT onset primary open angle glaucoma." Ophthalmic Genet. 18:109-118(1997).

RN [11] VARIANTS JOAG ARG-367 AND LEU-370.

RX MEDLINE: 98141135.

RA MICHELIS-RAUTENSTRAUSS K.G., MARDIN C.Y., BUDE W.M., LIEHR T.,

RA POLANSKY J., NGUYEN T., TIMMERMAN V., VAN BROECKHOVEN C.,

RA NAUMANN G.O.H., PREIFFER R.A., RAUTENSTRAUSS B.W.;

RT "Juvenile open angle glaucoma: fine mapping of the TIGR gene to

RT 1q24.3-q25.2 and mutation analysis." Hum. Genet. 102:103-106(1998).

RN [12] VARIANTS COAG ARG-367, AND VARIANT JOAG PHE-426.

RP MEDLINE: 98180724.

RA MANSERGH F.C., KENNA P.F., AYUSO C., KIANG A.-S., HUMPHRIES P.,

RA FARRAR G.J.;

RT "Novel mutations in the TIGR gene in early and late onset open angle

RT glaucoma." Hum. Mutat. 11:244-251(1998).

CC -1- FUNCTION: MAY PARTICIPATE IN THE OBSTRUCTION OF FLUID OUTFLOW IN

CC THE TRABECULAR MESHWORK.

CC -1- SUBCELLULAR LOCATION: LOCATED PREFERENTIALLY IN THE CILIARY

CC ROOTLET AND BASAL BODY OF THE CONNECTING CLIVUS OF PHOTORECEPTOR

CC CELLS, AND IN THE ROUGH ENDOPLASMIC RETICULUM, ALSO SECRETED.

CC -1- ALTERNATIVE PRODUCTS: THE SHORT FORM MAY BE THE PRODUCT OF AN

CC ALTERNATIVE INITIATION AT MET-15.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN LARGE AMOUNTS IN VARIOUS TYPES OF

CC MUSCLE, CILIARY BODY, PAPILLARY SPINCTER, SKELETAL MUSCLE, HEART

CC AND OTHER TISSUES. EXPRESSED PREDOMINANTLY IN THE RETINA. IN

CC NORMAL EYES, FOUND IN THE INNER VITREAL MESHWORK REGION AND THE

CC ANTERIOR PORTION OF THE MESHWORK. IN CONTRAST, IN MANY

CC GLAUCOMATOUS EYES, IT IS FOUND IN MORE REGIONS OF THE MESHWORK AND

CC APPEARED MORE INTENSIVELY THAN IN NORMAL EYES, REGARDLESS OF THE

CC TYPE OR CLINICAL SEVERITY OF GLAUCOMA.

CC -1- PTM: DIFFERENT ISOFORMS MAY ARISE BY POST-TRANSLATIONAL

CC MODIFICATIONS.

CC -1- DISEASE: DEFECTS IN MYOC ARE A CAUSE OF THE PRIMARY OPEN ANGLE

CC GLAUCOMA (POAG). POAG IS SUBDIVIDED INTO TWO DIFFERENT CATEGORIES

CC DEPENDING ON THE AGE OF ONSET: JUVENILE ONSET OPEN ANGLE GLAUCOMA

CC (JOAG OR GLCIA) AND CHRONIC OPEN ANGLE GLAUCOMA (COAG) WHICH IS

CC LATER IN ONSET. POAG IS A HIGHLY PREVALENT CAUSE OF IRREVERSIBLE

CC BLINDNESS. IT IS CHARACTERIZED BY COUPLING OF THE OPTIC DISK AND

CC ALTERATION OF THE VISUAL FIELD. ELEVATION OF INTRAOCULAR PRESSURE

CC IS OFTEN PRESENT AND IS A MAJOR RISK FACTOR. THE DISEASE IS

CC PAINLESS AND OFTEN DIAGNOSED AT A LATE STAGE, WHEN VISUAL FIELD

CC DEFECTS ARE SEVERE.

CC -1- SIMILARITY: BELONGS TO THE OLFACTOMEDIN FAMILY.

CC -1- This SWISS-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).

DR EMBL: U85257; AAC52051.1; -

DR EMBL: AF001620; AAC51725.1; -

DR EMBL: D88214; BAA23531.1; -

DR EMBL: 297171; CAB09899.1; JOINED.

DR EMBL: 297177; CAB09899.1; JOINED.

DR EMBL: 297174; CAB09899.1; JOINED.

DR EMBL: 298750; CAB11430.1; -

DR EMBL: AF049793; AAC14264.1; -

DR EMBL: AF049791; AAC14264.1; JOINED.

DR EMBL: AF049792; AAC14264.1; JOINED.

DR HSSP: P01100; IPOS.

DR MIM: 601652; -

DR MIM: 137750; -

KW Signal: Disease mutation; Glycoprotein; Alternative initiation;

KW Coiled coil.

FT SIGNAL 1 32 POTENTIAL.

FT CHAIN 33 504 MYOCILIN.

FT DOMAIN 74 184 COILED COIL (POTENTIAL).

FT CAROHD 310 486 OLFACTOMEDIN-LIKE.

FT SITE 502 504 POTENTIAL.

FT VARIANT 246 246 MICROBODY TARGETING SIGNAL (POTENTIAL).

FT VARIANT 337 337

FT VARIANT 364 364

FT VARIANT 367 367

FT VARIANT 370 370

FT VARIANT 426 426

FT VARIANT 437 437

FT VARIANT 477 477

FT VARIANT 480 480

FT VARIANT 499 499

FT SEQUENCE 504 AA; 56972 MW; 1919A632 CRC32;

QY 2 TOKIYRISGVGID 15

Db 281 TOETTRIDYVGT 294

RESULT 8

HMGL\_BOVIN STANDARD: PRT: 140 AA.

AC Q29448;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYDROXYMETHYLGITARTYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL) (3-

DE HYDROX-3-METHYLGITARTYL-COA LYASE) (FRAGMENT).

GN HMGL.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;

OC Bovinae; Bos.

CC [1]

CC SEQUENCE FROM N.A.

CC JI S., KUSE J., SPURLOCK M.E.;

CC Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROX-3-METHYLGITARTYL-COA -

CC ACETYL-COA + ACETOACETATE.

Query Match 52.7%; Score 39; DB 1; Length 504;

Best Local Similarity 57.1%; Pred. No. 14;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```

CC -1- PATHWAY: FINAL STEP OF KETOGENESIS AND LECICINE CATABOLISM.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U41409; AAA86757.1; -.
DR PROSITE: PS01062; HMG_COA_LYASE; 1.
KW Lyase; Mitochondrion.
FT ACT_SITE 1 1 BY SIMILARITY.
FT NON_TER 1 1
SO SEQUENCE 140 AA; 14409 MW; 08C3C0C3 CRC32;

Query Match      51.4%; Score 38; DB 1; Length 140;
Best Local Similarity 42.9%; Pred. No. 6.1;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATOKTTRISGVGI 14
   11: 1:1:1:1:1
Db 92 ATEDLVYMLAGLGI 105

RESULT 9
HMGCL_CHICK 9
HMGCL_CHICK STANDARD: PRT: 298 AA.
AC P35915;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYDROXYMETHYLGLOUTARYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL) (3-
DE HYDROXY-3-METHYLGLOUTARATE-COA LYASE).
GN HMGCL.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]
RP SEQUENCE OF 10-298 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-LIVER;
RX MITCHELL G.A., ROBERT M.-F., HRUZ P.W., WANG S., FONTAINE G.,
RA BENKE C.E., MENDE-MUELLER L.M., SCHAPPERT K., LEE C., GIBSON K.M.,
RA MIOTOKO H.M.;
RT "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL). Cloning of human
RT and chicken liver HL cDNAs and characterization of a mutation causing
RT human HL deficiency."
RL J. Biol. Chem. 268:4376-4381(1993).
CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLOUTARYL-COA -
CC ACETYL-COA + ACETOACETATE.
CC -1- PATHWAY: FINAL STEP OF KETOGENESIS AND LECICINE CATABOLISM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L07034; AAA92728.1; ALT_INIT.
DR PIR: B45470; B45470.
DR PFAM: PF00682; HMGCL-like; 1.
DR PROSITE: PS01062; HMG_COA_LYASE; 1.

```

```

KW Lyase; Mitochondrion.
FT ACT_SITE 239 239 BY SIMILARITY.
FT NON_TER 239 239
SO SEQUENCE 298 AA; 31436 MW; A2EE0702 CRC32;

Query Match      51.4%; Score 38; DB 1; Length 298;
Best Local Similarity 42.9%; Pred. No. 13;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATOKTTRISGVGI 14
   11: 1:1:1:1:1
Db 250 ATEDLVYMLAGLGI 263

RESULT 10
HMGCL_MOUSE 10
HMGCL_MOUSE STANDARD: PRT: 325 AA.
AC P38060;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYDROXYMETHYLGLOUTARYL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)
DE (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLOUTARATE-COA LYASE).
GN HMGCL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RX MEDLINE: 93364116.
RA WANG S., NADRAU J.H., DUNCAN A., ROBERT M.-F., FONTAINE G.,
RA SCHAPPERT K., JOHNSON K.R., ZIEKIEWICZ E., HRUZ P., MIOTOKO H.;
RT "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL): cloning and
RT characterization of a mouse liver HL cDNA and subchromosomal mapping
RT of the human and mouse Hl genes."
RL Mamm. Genome 4:382-387(1993).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN-129;
CC MEDLINE: 96207305.
CC WANG S.P., ROBERT M.-F., GIBSON K.M., WANDERS R.J.A., MITCHELL G.A.;
RT "3-hydroxy-3-methylglutaryl coA lyase (HL): mouse and human HL gene
RT (HMGCL) cloning and detection of large gene deletions in two
RT unrelated HL-deficient patients."
RL Genomics 33:99-104(1996).
CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLOUTARYL-COA -
CC ACETYL-COA + ACETOACETATE.
CC -1- PATHWAY: FINAL STEP OF KETOGENESIS AND LECICINE CATABOLISM.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.
CC -1- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S65036; AAB27965.1; -.
DR EMBL: U49878; AAB03107.1; JOINED.
DR EMBL: U49870; AAB03107.1; JOINED.
DR EMBL: U49871; AAB03107.1; JOINED.
DR EMBL: U49872; AAB03107.1; JOINED.
DR EMBL: U49873; AAB03107.1; JOINED.
DR EMBL: U49874; AAB03107.1; JOINED.
DR EMBL: U49875; AAB03107.1; JOINED.
DR EMBL: U49876; AAB03107.1; JOINED.
DR EMBL: U49877; AAB03107.1; JOINED.
DR MGI: MGI:96158; HMGCL.
DR PFAM: PF00682; HMGCL-like; 1.

```

DR PROSITE: PS01062; HMG\_COA\_LYASE: 1.  
 KM Lyase: Mitochondrion; Transit peptide.  
 FT TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).  
 FT CHAIN 28 325 HYDROXYMETHYLGLOUTARL-COA LYASE.  
 FT ACT\_SITE 266 266 BY SIMILARITY.  
 FT CONFLICT 62 63 ML -> IV (IN REF. 2).  
 FT CONFLICT 80 80 N -> K (IN REF. 2).  
 FT CONFLICT 231 231 G -> A (IN REF. 2).  
 FT CONFLICT 238 238 I -> Y (IN REF. 2).  
 SO SEQUENCE 325 AA; 34161 MW; 3446C502 CRC32.

Query Match 51.4%; Score 38; DB 1; Length 325;  
 Best Local Similarity 42.9%; Pred. No. 14;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 AFOKITRISGVGI 14  
 ||:|:|:|:|:|  
 Db 277 ATEDLVYMLNGLGI 290

RESULT 11  
 HMG\_LYASE  
 ID HMG\_LYASE STANDARD; PRT; 325 AA.  
 AC P97519;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE HYDROXYMETHYLGLOUTARL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)  
 DE (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLOUTARATE-COA LYASE).  
 GN HMGCL.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MISTAR; TISSUE-LIVER;  
 RA COLLINGFORD T.E., DOLPHIN C.T., CANEVARI L., CLARK J.B.;  
 RL Submitted (JEC-1996) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLOUTARL-COA -  
 ACETYL-COA + ACETOACETATE.  
 CC -1- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.  
 CC -1- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.  
 CC .....  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC .....  
 CC EMBL: Y10054; CAA71148.1; -  
 CC PFM: PF00682; HMG\_LYASE: 1.  
 DR PROSITE: PS01062; HMG\_COA\_LYASE: 1.  
 KM Lyase: Mitochondrion; Transit peptide.  
 FT TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).  
 FT CHAIN 28 325 HYDROXYMETHYLGLOUTARL-COA LYASE.  
 FT ACT\_SITE 266 266 BY SIMILARITY.  
 SO SEQUENCE 325 AA; 34192 MW; C68DE719 CRC32;

Query Match 51.4%; Score 38; DB 1; Length 325;  
 Best Local Similarity 42.9%; Pred. No. 14;  
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 1 AFOKITRISGVGI 14  
 ||:|:|:|:|:|  
 Db 277 ATEDLVYMLNGLGI 290

RESULT 12  
 TSP3\_MOUSE  
 ID TSP3\_MOUSE STANDARD; PRT; 956 AA.  
 AC 005895;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE THROMBOSPONDIN 3 PRECURSOR.  
 GN THBS3 OR TSP3.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-B6/CBA; TISSUE-LUNG;  
 RA MEDLINE: 93224149.  
 RA BORNSTEIN P., DEVARAYALU S., EDELHOFF S., DISTECHE C.M.;  
 RT "Isolation and characterization of the mouse thrombospondin 3 (Thbs3)  
 gene.";  
 RT Genomics 15:607-613(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE: 94117438.  
 RA QABAR A.N., LIN Z., WOLF F.W., O'SHEA K.S., LAWLER J., DIXIT V.M.;  
 RT "Thrombospondin 3 is a developmentally regulated heparin binding  
 protein.";  
 RT J. Biol. Chem. 269:1262-1269(1994).  
 RN [3]  
 RP SEQUENCE OF 517-956 FROM N.A.  
 RC STRAIN-BALB/C;  
 RA MEDLINE: 92291102.  
 RA VOS H.L., DEVARAYALU S., DE VRIES Y., BORNSTEIN P.;  
 RT "Thrombospondin 3 (Thbs3), a new member of the thrombospondin gene  
 family.";  
 RT J. Biol. Chem. 267:12192-12196(1992).  
 RN [4]  
 RP SEQUENCE OF 834-956 FROM N.A.  
 RC STRAIN-BALB/C;  
 RA VOS H.L., MCKINSTRUM-WILSON M., ROOD P.M.L., MAAS A.C.E.,  
 RA DUBHIG T., GENDLER S.J., BORNSTEIN P.;  
 RL Submitted (GEN-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND  
 CC CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,  
 CC LAMININ AND TYPE V COLLAGEN.  
 CC .....  
 CC -1- SUBUNIT: OLIGOMER, CROSS-LINKED BY DISULFIDE BONDS.  
 CC -1- TISSUE SPECIFICITY: BRAIN, LUNGS AND CARTILAGE.  
 CC -1- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 7 TYPE-3 TSP REPEATS (WHICH BIND CALCIUM).  
 CC .....  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC .....  
 CC EMBL: L04302; AAA40497.1; -  
 CC EMBL: L24434; AAA40430.1; -  
 CC EMBL: M86620; AAA40430.1; JOINED.  
 CC EMBL: M86611; AAA40430.1; JOINED.  
 CC EMBL: M86612; AAA40430.1; JOINED.  
 CC EMBL: M86613; AAA40430.1; JOINED.  
 CC EMBL: M86614; AAA40430.1; JOINED.  
 CC EMBL: M86615; AAA40430.1; JOINED.  
 CC EMBL: M86616; AAA40430.1; JOINED.  
 CC EMBL: M86617; AAA40430.1; JOINED.  
 CC EMBL: M86618; AAA40430.1; JOINED.  
 CC EMBL: M86619; AAA40430.1; JOINED.  
 CC EMBL: M86620; AAA40430.1; JOINED.  
 CC PIR: A46016; A46016.  
 CC PIR: A44124; A44124.





ID DSG2 HUMAN STANDARD; PRT: 1117 AA.

AC 014126;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE DESMOGLEIN 2 PRECURSOR (HDGC).

GN DSG2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-COLON CARCINOMA;

RA MEDLINE; 94192736.

RA SCHAEFER S., KOCH P.J., FRANK W.W.;

RT "Identification of the ubiquitous human desmoglein, Dsg2, and the expression catalogue of the desmoglein subfamily of desmosomal cadherins."

RT Exp. Cell Res. 211:391-399(1994).

RL [2]

RN SEQUENCE OF 777-1117 FROM N.A.

RP MEDLINE; 92037656.

RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMMERMANN R., FRANK W.W.;

RT "Complete amino acid sequence of the epidermal desmoglein precursor polypeptide and identification of a second type of desmoglein gene."

RL Eur. J. Cell Biol. 55:200-208(1991).

CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.

CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE FILAMENT'S MEDIATING CELL-CELL ADHESION.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.

CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOsome SUBFAMILY.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC CC

DR EMBL: Z26317; CA81226.1; -

DR HSP; P15116; INCI.

DR MIM: 125671; -

DR PFAM: PF00028; cadherin; 4.

DR PROSITE: PS00232; CADHERIN; 3.

KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; signal; Cytoskeleton; Calcium-binding.

CC SIGNAL 1 23

FT PROPEP 24 48

FT CHAIN 49 1117

FT DOMAIN 49 608

FT TRANSEM 609 633

FT DOMAIN 634 1117

FT REPEAT 49 159

FT REPEAT 160 272

FT REPEAT 273 387

FT REPEAT 388 502

FT REPEAT 503 617

FT REPEAT 618 732

FT REPEAT 733 847

FT REPEAT 848 962

FT REPEAT 963 1078

FT REPEAT 1079 1194

FT CARBOHYD 181 181

FT CARBOHYD 308 308

FT CARBOHYD 461 461

FT CARBOHYD 513 513

SO SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;

Query Match 50.0%; Score 37; DB 1; Length 1117;

Best Local Similarity 63.6%; Pred. No. 74;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 KITIRISGVI 14

DB 83 KITIKYTKGI 93

RESULT 15

ID HMGL HUMAN STANDARD; PRT: 325 AA.

AC P35914;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE HYDROXYMETHYLGLUTARYL-COA LYASE, MITOCHONDRIAL PRECURSOR (EC 4.1.3.4)

DE (HMG-COA LYASE) (HL) (3-HYDROXY-3-METHYLGLUTARATE-COA LYASE).

GN HMGL.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LIVER;

RA MEDLINE; 93179448.

RA MITCHELL G.A., ROBERT M.-F., HRUZ P.W., WANG S., FONTAINE G., BEHNE C.E., MENDE-MILLER L.M., SCHAPPERT K., LEE C., GIBSON K.M., MIZIORKO H.M.;

RT "3-hydroxy-3-methylglutaryl coenzyme A lyase (HL): Cloning of human and chicken liver HL cDNAs and characterization of a mutation causing human HL deficiency."

RT J. Biol. Chem. 268:4376-4381(1993).

RL [2]

RN SEQUENCE OF 21-325 FROM N.A.

RP MEDLINE; 96207305.

RA WANG S.P., ROBERT M.-F., GIBSON K.M., WANDERS R.J.A., MITCHELL G.A.;

RT "3-Hydroxy-3-methylglutaryl CoA lyase (HL): mouse and human HL gene (HMGL) cloning and detection of large gene deletions in two unrelated HL-deficient patients."

RL Hum. Genomics 33:99-104(1996).

RN [3]

RP VARIANT HMG ARG-233.

RA MEDLINE; 96394619.

RA ROBERTS J., MITCHELL G.A., MIZIORKO H.M.;

RT "Modeling of mutational responsible for human 3-hydroxy-3-methylglutaryl-CoA lyase deficiency implicates histidine-233 as an active site residue."

RL J. Biol. Chem. 271:24604-24609(1996).

RN [4]

RP VARIANTS HMG GLN-41; GLU-42; GLY-42 AND HIS-42.

RA MEDLINE; 96130531.

RA MITCHELL G.A., OZAND P.T., ROBERT M.-F., ASHMARINA L., ROBERTS J., GIBSON K.M., WANDERS R.J., WANG S., CHEVALIER I., PLOEHL E., MIZIORKO H.;

RT "HMG CoA lyase deficiency: identification of five causal point mutations in codons 41 and 42, including a frequent Saudi Arabian mutation, R41Q."

RL Am. J. Hum. Genet. 62:295-300(1998).

CC -1- CATALYTIC ACTIVITY: (S)-3-HYDROXY-3-METHYLGLUTARYL-COA - ACETYL-COA + ACETOACETATE.

CC -1- PATHWAY: FINAL STEP OF KETOGENESIS AND LEUCINE CATABOLISM.

CC -1- SUBUNIT: HOMODIMER.

CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL MATRIX.

CC -1- TISSUE SPECIFICITY: FIBROBLASTS, LIVER AND LYMPHOBLASTS.

CC -1- DISEASE: DEFECTS IN HMGL ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE DISEASE KNOWN AS HYDROXYMETHYLGLUTARICACIDURIA (HMG) WHICH CAN LEAD TO HIPPOGLOMIA AND COMA.

CC -1- SIMILARITY: BELONGS TO THE HMG-COA LYASE FAMILY.

CC CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL: L07033; AAA2733.1; -.  
DR EMBL: U49719; AAB19099.1; -.  
DR EMBL: U49712; AAB19099.1; JOINED.  
DR EMBL: U49713; AAB19099.1; JOINED.  
DR EMBL: U49714; AAB19099.1; JOINED.  
DR EMBL: U49715; AAB19099.1; JOINED.  
DR EMBL: U49716; AAB19099.1; JOINED.  
DR EMBL: U49717; AAB19099.1; JOINED.  
DR EMBL: U49718; AAB19099.1; JOINED.  
DR PIR: A45470; A45470.  
DR MIM: 246450; -.  
DR PFAM: PF00682; HMGL-1like; 1.  
DR PROSITE: PS01062; HMG\_COA\_LYASE; 1.  
KV Lyase: Mitochondrion: Transit peptide: Disease mutation.  
FT TRANSIT 1 27 MITOCHONDRION (BY SIMILARITY).  
FT CHAIN 28 325 HYDROXYMETHYLGLUTARYL-COA LYASE.  
FT ACT\_SITE 266 266 BY SIMILARITY.  
FT VARIANT 41 41 R -> Q (IN HMG).  
FT VARIANT 42 42 /FTID-VAR\_003744.  
FT VARIANT 42 42 D -> E (IN HMG).  
FT VARIANT 42 42 /FTID-VAR\_003745.  
FT VARIANT 42 42 D -> G (IN HMG).  
FT VARIANT 42 42 /FTID-VAR\_003746.  
FT VARIANT 42 42 D -> H (IN HMG).  
FT VARIANT 70 70 /FTID-VAR\_003747.  
FT VARIANT 70 70 V -> L (IN HMG).  
FT VARIANT 233 233 /FTID-VAR\_003748.  
FT VARIANT 233 233 H -> R (IN HMG).  
FT CONFLICT 243 243 /FTID-VAR\_003749.  
FT SEQUENCE 325 AA; 34390 MW; DCDBABC8 CRC32; T -> A (IN REF. 2).

Query Match 50.0%; Score 37; DB 1; Length 325;  
Best Local Similarity 42.9%; Pred. No. 21;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AFOKITVRSIGVGI 14  
||:|:|:|:|:|:|  
DB 277 ATEDLVYMEGLGI 290

Search completed: January 7, 2000, 13:25:35  
Job time: 387 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OW protein - protein search, using sw model

Run on: January 12, 2000, 23:12:17 : Search time 59.21 seconds  
(without alignments)  
15,518 Million cell updates/sec

Title: US-08-991-628-1

Perfect score: 74

Sequence: 1 AOKITRISGVGID 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL\_11.\*

Word size: 0

Number of hits that pass the threshold : 199794

1: sp\_archae:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	65	87.8	993	11	035902	035902 mus musculu
2	40	54.1	2023	5	096542	096542 drosophila
3	39	52.7	3380	5	024292	024292 drosophila
4	39	52.7	441	10	022282	022282 aradidopsi
5	38	51.4	147	1	026829	026829 mechanobact
6	38	51.4	331	2	066968	066968 aquiflex aeo
7	38	51.4	303	2	P95639	P95639 rhodospirill
8	36	51.4	177	3	012036	012036 saccharomyc
9	38	51.4	2269	5	077360	077360 plasmodium
10	38	51.4	287	5	020481	020481 caenorhabdi
11	38	51.4	82	7	077862	077862 oreochromis
12	38	51.4	1177	10	023377	023377 aradidopsi
13	38	51.4	433	10	081037	081037 aradidopsi
14	37	50.0	402	2	083781	083781 treponema p
15	37	50.0	2408	4	092566	092566 homo sapien
16	37	50.0	1161	5	004901	004901 halocynthia
17	37	50.0	63	7	046802	046802 poeciliopsi
18	37	50.0	1363	12	066290	066290 human coron
19	37	50.0	1363	12	066291	066291 human coron
20	36	48.6	296	2	031431	031431 bacillus su
21	36	48.6	482	2	059326	059326 clostridium
22	36	48.6	507	2	085888	085888 spirillum
23	36	48.6	381	2	047594	047594 escherichia
24	36	48.6	191	4	095896	095896 homo sapien
25	36	48.6	839	5	062002	062002 branchiosto

## ALIGNMENTS

26	36	48.6	3097	5	015943	015943 drosophila
27	36	48.6	76	7	046796	046796 poeciliopsi
28	36	48.6	82	7	031538	031538 pseudotroph
29	36	48.6	82	7	031536	031536 pseudotroph
30	36	48.6	82	7	031537	031537 pseudotroph
31	36	48.6	1065	10	048948	048948 aradidopsi
32	36	48.6	138	12	090818	090818 human immun
33	36	48.6	717	13	091353	091353 gallus gall
34	35	47.3	294	1	054288	054288 sulfobolus
35	35	47.3	855	1	057852	057852 pyrococcus
36	35	47.3	420	2	068332	068332 pseudomonas
37	35	47.3	149	2	048608	048608 lactococcus
38	35	47.3	780	2	P77846	P77846 clostridium
39	35	47.3	362	2	069445	069445 methylococc
40	35	47.3	339	2	085193	085193 streptococ
41	35	47.3	331	2	092950	092950 bacillus sp
42	35	47.3	878	4	015855	015855 homo sapien
43	35	47.3	1223	5	044327	044327 caenorhabdi
44	35	47.3	432	5	061820	061820 caenorhabdi
45	35	47.3	572	5	023919	023919 dictyostell

RESULT 1  
ID 035902 PRELIMINARY; PRT; 993 AA.  
AC 035902;  
DT 01-JAN-1998 (TRENKREL. 05, Created)  
DT 01-JAN-1998 (TRENKREL. 05, Last sequence update)  
DT 01-MAY-1999 (TRENKREL. 10, Last annotation update)  
DE DESMOGLEIN 3 (FRAGMENT).  
GN DSG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
CC Eukarya; Rodentia; Sciurognathi; Muridae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RA ISHIKAWA H., LI K., UIRTO J.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; U86016; AAB5091.1; -.  
DR PFM; PFM0028; cadherin; 4.  
DR PROSITE; PS00232; CADHERIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON-TER 993  
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 87.8%; Score 65; DB 11; Length 993;  
Best Local Similarity 100.0%; Pred. No. 0.0025;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07 3 OKITRISGVGID 15  
Db 81 OKITRISGVGID 93  
RESULT 2  
ID 096542 PRELIMINARY; PRT; 2023 AA.  
AC 096542;  
DT 01-MAY-1999 (TRENKREL. 10, Created)  
DT 01-MAY-1999 (TRENKREL. 10, Last sequence update)  
DT 01-MAY-1999 (TRENKREL. 10, Last annotation update)  
DE ENHANCER OF POLYCOMB.  
GN E.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
CC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-IMAGINAL DISC;  
 RX MEDLINE: 98407961.  
 RA STANNUNIS K., BERGER J., ROSE C., SINCLAIR D.A.R., RANDAZZO F.,  
 RA BROCK H.W.;  
 RT "The Enhancer of Polycomb gene of Drosophila encodes a chromatin  
 RT protein conserved in yeast and mammals."  
 RL Development 125:4055-4066(1998).  
 DR EMBL: AF079764; AAC64271.1; -;  
 SQ SEQUENCE 2023 AA; 220680 MW; 1D7D99F6 CRC32;

Query Match 54.1%; Score 40; DB 5; Length 2023;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATOKITRISGVG 14  
 DB 1453 ASOKVYAIQKAGV 1466

RESULT 3  
 ID 024292 PRELIMINARY; PRT; 3380 AA.  
 AC 024292;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)  
 DE ADHERIN.  
 GN DACHSOS.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 95324813.  
 RA CLARK H.F., BRENTNUP D., SCHNEITZ K., BIBBER A., GOODMAN C., NOLL M.;  
 RT "Dachsous encodes a member of the cadherin superfamily that controls  
 RT imaginal disc morphogenesis in Drosophila."  
 RL Genes Dev. 9:1530-1542(1995).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: L08811; AAA9329.1; -;  
 DR FLYBASE: FBgn0000497; ds.  
 DR PRAM: PF000028; cadherin; 25.  
 DR PROSITE: PS00232; CADHERIN; 18.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 3380 AA; 366356 MW; 7D83FAC3 CRC32;

Query Match 52.7%; Score 39; DB 5; Length 3380;  
 Best Local Similarity 58.3%; Pred. No. 4.4e+02;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 4 KITRISGVG 15  
 DB 1017 RITRYIVDAGVD 1028

RESULT 4  
 ID 022282 PRELIMINARY; PRT; 441 AA.  
 AC 022282;  
 DT 01-JAN-1998 (TEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-JUN-1998 (TEMBLrel. 06, Last annotation update)  
 DE T517.1 OR F17A14.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
 RA SYKES S.M., KAUL S., MASON T.M., KERLAJAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RL Submitted (Oct-1997) to the EMBL/Genbank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA ROUNSLEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
 RA SYKES S.M., KAUL S., MASON T.M., KERLAJAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RL Submitted (Dec-1997) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AC003000; AAB67120.1; -;  
 DR EMBL: AC003674; AAB97126.1; -;  
 SQ SEQUENCE 441 AA; 47873 MW; F462C421 CRC32;

Query Match 52.7%; Score 39; DB 10; Length 441;  
 Best Local Similarity 54.5%; Pred. No. 44;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 OKITRISGVG 13  
 DB 368 OKLYRVNGAG 378

RESULT 5  
 ID 026829 PRELIMINARY; PRT; 147 AA.  
 AC 026829;  
 DT 01-JAN-1998 (TEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TEMBLrel. 07, Last annotation update)  
 DE HYPOTHETICAL 15.6 KD PROTEIN.  
 GN MTH733.  
 OS Methanobacterium thermoautotrophicum.  
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;  
 OC Methanobacterium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-DELTA H;  
 RX MEDLINE: 98037514.  
 RA SMITH D.R., DOUCETTE-STAMM L.A., DELOUGHERY C., LEE H.-M., DUBOIS J.,  
 RA ALDENDE T., BASHIRZADEH R., BLAKELY D., COOK R., GILBERT K.,  
 RA HARRISON D., HOANG L., KEAGLE P., LOWE W., POTIER B., QIU D.,  
 RA SPADAFORA R., VICARE R., WANG Y., WIERZBOWSKI J., GIBSON R.,  
 RA JIWANI N., CARUSO A., BUSH D., SAVER H., PATWELL D., PRABHAKAR S.,  
 RA MCDUGALL S., SHIMER G., GOYAL A., PIETROVSKI S., CHURCH G.M.,  
 RA DANIELS C.J., MAO J.-I., RICE P., NOLLING J., REEVE J.N.;  
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum  
 RT deltaH: functional analysis and comparative genomics."  
 RL J. Bacteriol. 179:7135-7155(1997).  
 DR EMBL: AE000852; AAB85238.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 147 AA; 15626 MW; 9029EB848 CRC32;

Query Match 51.4%; Score 38; DB 1; Length 147;  
 Best Local Similarity 61.5%; Pred. No. 19;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATOKITRISGVG 13  
 DB 135 ATYKYNRTGAGV 147

RESULT 6  
 ID 066968 PRELIMINARY; PRT; 331 AA.  
 AC 066968;

DT 01-AUG-1998 (Tremblrel. 07, Created)  
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE PHOSPHORIBOSYLTRANSFERASE CYCLO-LIGASE.  
 GN PRM.  
 OS Aquifex aeolicus.  
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RX MEDLINE; 98196666.  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUTAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RT "The complete genome of the hyperthermophilic bacterium Aquifex  
 aeolicus.";  
 RL Nature 392:353-358(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-VF5;  
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,  
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., AUTAY M., HUBER R.,  
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;  
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE000706; AAC06918.1; -;  
 DR PRM; PF00586; AIRS; 1.  
 KW Ligase.  
 SO SEQUENCE 331 AA; 36768 MW; 4C3D13B0 CRC32;

Query Match 51.4%; Score 38; DB 2; Length 331;  
 Best Local Similarity 63.6%; Pred. No. 48;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Y 5 ITYRISGVGD 15  
 D 2 VTRISAGVDID 12

RESULT 7  
 P95639 PRELIMINARY; PRT; 303 AA.  
 AC P95639;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE HYDROXYMETHYLGUTARYL-COA LYASE (EC 4.1.3.4) (HMG-COA LYASE) (HL)  
 DE (3-HYDROXY-3-METHYLGUTARATE-COA LYASE).  
 GN HMG.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;  
 CC Rhodospirillum.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S1;  
 RX BALTSCHIEFSKY M., BROSCHE M., HOUTMAN T., LUNDYK L., NYREN P.,  
 RA MALESCHIEFSKY M., SEVERIN A., STRID A.;  
 RA SAKAI-NORE Y., SEVERIN A., STRID A.;  
 RT "A 3-hydroxy-3-methylglutaryl-CoA lyase gene in the photosynthetic  
 bacterium Rhodospirillum rubrum.";  
 RL Biochim. Biophys. Acta 1337:113-122(1997).  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S1;  
 RX BALTSCHIEFSKY M., BROSCHE M., HOUTMAN T., LUNDYK L., NYREN P.,  
 RA MALESCHIEFSKY M., SEVERIN A., STRID A.;  
 RA SAKAI-NORE Y., SEVERIN A., STRID A.;  
 RT "A 3-hydroxy-3-methylglutaryl-CoA lyase gene in the photosynthetic  
 bacterium Rhodospirillum rubrum.";  
 RL Biochim. Biophys. Acta 1337:113-122(1997).  
 CC (S)-3-HYDROXY-3-METHYLGUTARYL-COA - ACETYL-COA + ACETOACETATE.  
 CC -1- SIMILARITY: TO OTHER SPECIES HMG-COA LYASE.  
 DR EMBL; U41280; AAB50182.1; -;  
 DR PRM; PF00682; HMG\_Like; 1.  
 DR PROSITE; PS01062; HMG\_COA\_LYASE; 1.  
 KW Ligase.  
 FT ACT\_SITE 239 239 BY SIMILARITY.  
 SO SEQUENCE 303 AA; 31138 MW; 44C2DA4B CRC32;

Query Match 51.4%; Score 38; DB 2; Length 303;

Best Local Similarity 42.9%; Pred. No. 43;  
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 Y 1 ATOKITRISGVGI 14  
 D 250 ATEDVYVYMDGMGI 263

RESULT 8  
 Q12036 PRELIMINARY; PRT; 177 AA.  
 AC Q12036;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-1998 (Tremblrel. 06, Last annotation update)  
 DE P2604 PROTEIN (ORF YPL148C).  
 GN PPT2.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;  
 OC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AB972;  
 RA PURNELLE B., COMBLET S., COSTER F., NAVEAU F., GOFFEAU A.;  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MIPPS;  
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STUBBLE H.P., MEIER S., WAGNER C., SCHWEIZER E.;  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; X96770; CAA65545.1; -;  
 DR EMBL; X96770; CAA65545.1; -;  
 DR EMBL; 273504; CAA97853.1; -;  
 DR EMBL; Y16253; CAA76138.1; -;  
 KW Transferase; Phosphopentetheine.  
 SO SEQUENCE 177 AA; 20398 MW; 2A4460D2 CRC32;

Query Match 51.4%; Score 38; DB 3; Length 177;  
 Best Local Similarity 46.7%; Pred. No. 23;  
 Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;  
 Y 1 ATOKITRISGVGD 15  
 D 8 ASRNIGRIAGVD 22

RESULT 9  
 Q77360 PRELIMINARY; PRT; 2269 AA.  
 AC Q77360;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)  
 DE MAL3P4.25 PROTEIN.  
 GN MAL3P4.25.  
 OS Plasmodium falciparum.  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-3D7;  
 RA HAMILIN N., LAWSON D., BARRELL B.;  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL008970; CAA15615.1; -;  
 DR PRM; PF00271; helicase\_C; 1.  
 SO SEQUENCE 2269 AA; 267233 MW; 11F1C8E0 CRC32;

Query Match 51.4%; Score 38; DB 5; Length 2269;  
 Best Local Similarity 48.9%; Pred. No. 4,3e+02;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;



RA ADAMS M.D., SOMERVILLE C.R., VENTER J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC F13C20 genomic sequence."  
 RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC005168; AAC32247.1;  
 DR PFW; PF00682; HMGL-1; 1.  
 KW Lyase.  
 SO SEQUENCE 433 AA; 46385 MW; A449D9B9 CRC32;

Query Match 51.4%; Score 38; DB 10; Length 433;  
 Best Local Similarity 35.7%; Pred. No. 65;  
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 ATOKTIRISGVI 14  
 ||:|:|:|:  
 DB 377 ATEDVYMLNGIGV 390

## RESULT 14

083781 PRELIMINARY; PRT; 402 AA.

AC 083781;  
 DT 01-NOV-1998 (TREMBLrel. 08. Created)  
 DT 01-NOV-1998 (TREMBLrel. 08. Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10. Last annotation update)  
 DE HYPOTHETICAL 44.6 KD PROTEIN.

OS Treponema pallidum.  
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.  
 RN [1]

RP SEQUENCE FROM N.A.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DOOSON R., GWINN M., HICKER E.K., CLAYTON R., KETCHUM K.A.,  
 RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RT "Complete genome sequence of Treponema pallidum, the syphilis  
 spirochete."  
 RL Science 281:375-388(1998).

RT [2]  
 RL SEQUENCE FROM N.A.

RA FRASER C.M., NORRIS S.J., WEINSTOCK G.M., WHITE O., SUTTON G.G.,  
 RA DOOSON R., GWINN M., HICKER E.K., CLAYTON R., KETCHUM K.A.,  
 RA SODERGEN E., HARDHAM J.M., MCLEOD M.P., SALZBERG S., PETERSON J.,  
 RA KHALAK H., RICHARDSON D., HOWELL J.K., CHIDAMBARAM M., UTTERBACK T.,  
 RA MCDONALD L., ARTIACH P., BOWMAN C., COTTON M.D., FUJII C., GARLAND S.,  
 RA HATCH B., HORST K., ROBERTS K., WATTHEY L., WEIDMAN J., SMITH H.O.,  
 RA VENTER J.C.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF001251; AAC65775.1;  
 DR PFW; PF01368; DH; 1.  
 KW Hypothetical protein.

SO SEQUENCE 402 AA; 44577 MW; C237C707 CRC32;

Query Match 50.0%; Score 37; DB 2; Length 402;  
 Best Local Similarity 57.1%; Pred. No. 90;  
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 ATOKTIRISGVI 14  
 ||:|:|:|:  
 DB 309 ATSNIAESISGIGI 322

## RESULT 15

092566 PRELIMINARY; PRT; 2408 AA.

AC 092566;  
 DT 01-FEB-1997 (TREMBLrel. 02. Created)  
 DT 01-FEB-1997 (TREMBLrel. 02. Last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10. Last annotation update)  
 DE MYELOBLAST KIAA0279 (FRAGMENT).  
 GN KIAA0279.

OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 CC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE-BRAIN.

RA MEDLINE; 97191544.

RA NAGASE T., SEKI N., ISHIKAWA K., OHIRA M., KAWARABAYASHI Y., OHARA O.,  
 RA TANAKA A., KOTANI H., MIYAJIMA N., NOMURA N.;  
 RT "Prediction of the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by  
 RT analysis of cDNA clones from cell line KG-1 and brain."  
 RL DNA Res. 3:321-329(1998).

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

DR EMBL; D87469; BA13407.1;  
 DR PFW; PF000028; cadherin; 5.

DR PFW; PF00008; EGF; 6.

DR PFW; PF00054; laminin; G; 1.

DR PROSITE; PS00232; CDHERIN; 4.

KW Glycoprotein; Cell adhesion; Transmembrane; Calcium-binding; Repeat.  
 FT NON-TER 1  
 SO SEQUENCE 2408 AA; 261740 MW; CDBA2001 CRC32;

Query Match 50.0%; Score 37; DB 4; Length 2408;  
 Best Local Similarity 50.0%; Pred. No. 6.9e+02;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 4 KITRISGVIGD 15  
 ::||:|:|:|:  
 DB 24 RLEIRLAGVGD 35

Search completed: January 12, 2000, 23:15:18  
 Job time: 181 sec





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:34 ; Search time 51.33 Seconds  
(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-2  
Perfect score: 78  
Sequence: 1 FGIFVVDKMTGDIINI 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36.\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	78	100.0	999	1	R30742	Human pemphigus vu
2	78	100.0	30	1	R33960	Peptide combining
3	78	100.0	614	1	W07908	Pemphigus vulgaris
4	78	100.0	15	1	W04842	Self epitope of de
5	64	82.1	19	1	R33961	Peptide combining
6	64	82.1	21	1	R33962	Peptide combining
7	56	71.8	560	1	W13009	Segment of desmoso
8	55	70.5	778	1	W15489	Pemphigus foliaceu
9	48	61.5	17	1	R33963	Peptide combining
10	42	53.8	713	1	W13136	Full length human
11	42	53.8	713	1	W25638	Human cadherin-13.
12	41	52.6	794	1	W13135	Putative human cad
13	41	52.6	794	1	W25637	Human cadherin-12.
14	39	50.0	1091	1	R28822	Alpha 6B integrin
15	39	50.0	1073	1	R55272	Human subunit of 1
16	38	48.7	797	1	R86866	Human protocadheri
17	38	48.7	433	1	W41732	Arabidopsis chloro
18	38	48.7	341	1	W53939	L-lactis MG1363 or
19	37	47.4	863	1	R22357	Carrot aspartokina
20	37	47.4	878	1	R55060	Sequence of human
21	37	47.4	499	1	R60609	Tobamovirus replic
22	37	47.4	845	1	R70065	Hepatitis B virus
23	37	47.4	878	1	R85487	Human E-cadherin p
24	37	47.4	616	1	R86867	Rat protocadherin
25	37	47.4	982	1	W20980	H. pylori secreted
26	37	47.4	863	1	W75440	Carrot aspartokina
27	37	47.4	978	1	W10992	H. pylori ORF 05ep
28	36	46.2	566	1	R13228	Entholucanase enco
29	36	46.2	1203	1	R58911	Product of alterna
30	36	46.2	1026	1	R58906	Human protocadheri
31	36	46.2	515	1	R69505	Aspergillus sp. re
32	36	46.2	511	1	R69506	Aspergillus sp. re
33	36	46.2	515	1	R69504	Aspergillus sp. re
34	36	46.2	883	1	R69627	Bacteriophage T7 R
35	36	46.2	1203	1	R87152	Alternatively sp11
36	36	46.2	1026	1	R87146	Protocadherin clon
37	36	46.2	488	1	R85875	WD-40 domain-contg
38	36	46.2	31	1	R85793	Peptide r1 from th
39	36	46.2	954	1	W19752	Yeast inhibitor of

40	36	46.2	20	1	W40250	Human wild-type E-
41	36	46.2	52	1	Y11246	Streptococcus pneu
42	36	46.2	883	1	Y01322	Wild-type T7 RNA p
43	36	46.2	883	1	Y01353	Wild-type T7 RNA p
44	36	46.2	883	1	Y01286	Wild-type RNA poly
45	35	44.9	712	1	R27823	Sequence encoded b

## ALIGNMENTS

RESULT 1	
ID R30742	
AC R30742;	
DT 14-JUN-1993 (first entry)	
DE Human pemphigus vulgaris 130kd antigen.	
KW Pemphigus vulgaris; skin disease; autoantibodies;	
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.	
OS Homo sapiens.	
PN US7798918-A.	
PD 15-DEC-1992.	
PF 27-NOV-1991; 798918.	
PR 27-NOV-1991; US-798918.	
PA (USSH.) US DEPT HEALTH & HUMAN SERVICE.	
PI Amagai M, Klaus-Kovtun V, Stanley JR.	
DR WPI: 93-067436/08.	
DR N-PSDB: Q35992.	
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for	
PT diagnostic and therapeutic uses	
PS Disclosure, Fig 7; 50pp; English.	
CC This sequence is the pemphigus vulgaris 130kd antigen. The protein	
CC and its encoding DNA may be used in the diagnosis and treatment of	
CC pemphigus vulgaris. It is thought that the antigen may be a cell	
CC adhesion molecule.	
SO Sequence 999 AA;	
Query Match	100.0%; Score 78; DB 1; Length 999;
Best Local Similarity	100.0%; Pred. No. 1e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 FGIFVVDKMTGDIINI 15	
DB 97 FGIFVVDKMTGDIINI 111	
RESULT 2	
ID R93960	
AC R93960;	
DT 05-JUL-1996 (first entry)	
DE Peptide combining with anti-interepidermal cellular antibody.	
KW anti-interepidermal cellular antibody; autoantibody; adsorbent.	
OS Synthetic.	
PN J07305893-A.	
PD 28-NOV-1995.	
PF 18-MAY-1994; 129556.	
PR 18-MAY-1994; JP-129556.	
PA (KURS.) KURARAY CO LTD.	
DR WPI: 96-045392/05.	
PT Anti-interepidermal cellular antibody-combining peptide - which can	
PT be immobilised on column to form adsorbent useful for treating	
PT diseases related to the antibody	
PS Claim 1; Page 2; 7pp; Japanese.	
CC A new peptide is disclosed which contains at least 5 contiguous amino	
CC acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe	
CC Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg	
CC Ala Glu (the present sequence), the peptide not containing more than	
CC 50 residues. The peptide combines with anti-interepidermal cellular	
CC antibody. It can be immobilised on a carrier to prepare an absorbent	
CC useful for the treatment of diseases related to anti-interepidermal	
CC cellular antibody.	

SQ Sequence 30 AA;

Query Match  
Best Local Similarity 100.0%; Score 78; DB 1; Length 30;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDIINI 15  
|||||

DB 8 FGIFVVDKNTGDIINI 22

RESULT 3

W07908  
W07908 standard; peptide; 614 AA.

AC W07908;  
DT 29-JAN-1997 (first entry)  
DE Pemphigus vulgaris antigen protein extracellular region.  
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;  
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;  
KW dermatology.  
OS Homo sapiens.  
PN J0818540-A.  
PD 23-JUL-1996.  
PF 30-JUN-1995; 165632.  
PA (NISHU) NISHUKAWA T.  
DR WPI; 96-388562/39.  
PT Fused protein recognised by pemphigus vulgaris auto-antibody -  
PS useful to treat and diagnose pemphigus vulgaris  
PS Claim 1: Page 7-9; 9pp; Japanese  
CC W07908 represents the human pemphigus vulgaris (PV) antigen  
CC extracellular region. The PV antigen is produced in patients with  
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare  
CC relapsing disease causing subepithelial, intra-epidermal bullae  
CC (vesicles) of the skin and mucous membranes, which is fatal if  
CC untreated. The PV antigen was fused to a human IgG1 hinge region  
CC and the resulting fusion protein is useful to treat or diagnose  
CC pemphigus vulgaris.  
SQ Sequence 614 AA;

Query Match  
Best Local Similarity 100.0%; Score 78; DB 1; Length 614;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDIINI 15  
|||||

DB 96 FGIFVVDKNTGDIINI 110

RESULT 4

W04842  
W04842 standard; peptide; 15 AA.

AC W04842;  
DT 18-FEB-1997 (first entry)  
DE Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerisation; self-epitope; antigen; autoimmune disease;  
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphomannomutase;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW Influenza; haemagglutinin; reovirus; sigma protein.  
OS Homo sapiens.  
PN W09627387-A1.  
PD 12-SEP-1996.  
PF 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PA (HARD) HARVARD COLLEGE.  
PT Strominger JL; Wucherpfennig KW;  
DR WPI; 96-453218/42.  
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for

PT Identification of other self and non-self antigens implicated in  
PT auto-immune disease  
PS Claim 1; Page 38; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise  
CC either an isolated human non-collagen or non-myosin basic protein  
CC (MBP) polypeptide which is capable of tolerising an individual to an  
CC autoantigen; or an isolated human pathogen polypeptide capable of  
CC tolerising an individual to that polypeptide. In both cases, the  
CC polypeptide (whether self or non-self) includes an amino acid  
CC sequence corresponding to a sequence motif for a MHC class II  
CC protein, such as HLA-DR, which is associated with a human autoimmune  
CC disease and which binds to the polypeptide to activate autoreactive  
CC T-cells in individuals with the autoimmune disease. This peptide is  
CC derived from the human desmoglein 3 protein (amino acids 97-111)  
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
CC derived from the human desmoglein protein are described in W04841-47.  
SQ Sequence 15 AA;

Query Match  
Best Local Similarity 100.0%; Score 78; DB 1; Length 15;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDIINI 15  
|||||

DB 1 FGIFVVDKNTGDIINI 15

RESULT 5

R93961  
R93961 standard; peptide; 19 AA.

AC R93961;  
DT 05-JUL-1996 (first entry)  
DE Peptide combining with anti-interdermal cellular antibody.  
KW Anti-interdermal cellular antibody; autoantibody; adsorbent.  
OS Synthetic.

PN J07309893-A.  
PD 28-NOV-1995.  
PF 18-MAY-1994; 129556.  
PR 18-MAY-1994; JP-129556.  
PA (KURS) KURARAY CO LTD.  
DR WPI; 96-045392/05.

PT Anti-interdermal cellular antibody-combining peptide - which can  
PT be immobilised on column to form adsorbent useful for treating  
PT diseases related to the antibody  
PS Example 1; Page 4; 7pp; Japanese.  
CC New peptides are disclosed which contain at least 5 contiguous amino  
CC acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe  
CC Val Val Asp Lys Asn Thr Gly Asp Ile Asn Ile Thr Ala Ile Val Asp Arg  
CC Glu Glu (see R93960), the peptide not containing more than 50 residues.  
CC The peptide combines with anti-interdermal cellular antibody. It can  
CC be immobilised on a carrier to prepare an adsorbent useful for the  
CC treatment of diseases related to anti-interdermal cellular antibody.  
CC The present sequence is a specific example of the new peptides.

SQ Sequence 19 AA;

Query Match  
Best Local Similarity 82.1%; Score 64; DB 1; Length 19;  
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDI 12  
|||||

DB 8 FGIFVVDKNTGDI 19

RESULT 6

R93962  
R93962 standard; peptide; 21 AA.

AC R93962;  
DT 05-JUL-1996 (first entry)  
DE Peptide combining with anti-interdermal cellular antibody.  
KW anti-interdermal cellular antibody; autoantibody; adsorbent.

OS Synthetic.  
 PN J07309893-A.  
 PD 28-NOV-1995.  
 PF 18-MAY-1994; 129556.  
 PR 18-MAY-1994; JP-129556.  
 PA (KURS ) KURARAY CO LTD.  
 DR WPI; 96-045392/05.  
 PT Anti-Interdermal cellular antibody-combining peptide - which can  
 be immobilised on column to form adsorbent useful for treating  
 diseases related to the antibody.  
 PS Example 2; Page 5; 7pp; Japanese.  
 CC New peptides are disclosed which contain at least 5 contiguous amino  
 acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe  
 Val Asp Lys Asp Thr Gly Asp Ile Asp Ile Thr Ala Ile Val Asp Arg  
 CC Glu Glu (see R93960), the peptide not containing more than 50 residues.  
 CC The peptide combines with anti-interdermal cellular antibody. It can  
 be immobilised on a carrier to prepare an adsorbent useful for the  
 treatment of diseases related to anti-interdermal cellular antibody.  
 CC The present sequence is a specific example of the new peptides.  
 SO Sequence 21 AA;

Query Match 82.1%; Score 64; DB 1; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 3.7e-05;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGFVVDKNTGD 12  
 |||||  
 DB 10 GGFVVDKNTGD 21

RESULT 7  
 W13009  
 ID W13009 standard; protein; 560 AA.  
 AC W13009;  
 DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.  
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;  
 KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.  
 OS Homo sapiens.  
 PN DE19531033-A1.  
 PD 27-FEB-1997.  
 PF 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.  
 PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;  
 DR WPI; 97-146518/14.  
 PT Antibody reactive with part of desmosomal cadherin - exposed on  
 surface of epithelial or carcinoma cells, not bound to desmosomes,  
 PT useful for diagnosis and treatment of carcinoma micrometastases  
 PS Claim 7; Page 5; 8pp; German.  
 CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or  
 CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to  
 CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separately, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to  
 CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or  
 CC carcinoma.  
 SO Sequence 560 AA;

Query Match 71.8%; Score 56; DB 1; Length 560;  
 Best Local Similarity 60.0%; Pred. No. 0.039;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGFVVDKNTGDINI 15

DB 50 GGFVVDKNTGELNV 64  
 ||||| :|||:|

RESULT 8  
 W15489  
 ID W15489 standard; protein; 778 AA.  
 AC W15489;  
 DT 17-JUN-1997 (first entry)  
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.  
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
 KW extracellular region; antigen; hinge portion; skin;  
 KW dermatitis herpetiformis; fusion protein; detection; ss.  
 OS Chimeric - Homo sapiens.  
 FH Key  
 FT Location/Qualifiers  
 FT domain  
 FT 1..545  
 /note="Pemphigus foliaceus antigen protein"  
 PN J09077800-A.  
 PD 25-MAR-1997.  
 PF 12-SEP-1995; 260899.  
 PR 12-SEP-1995; JP-260899.  
 PA (NISH/) NISHIKAWA T.  
 DR WPI; 97-241758/22.  
 DR P-PSDB: T66428.  
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked  
 through the hinge region used to treat pemphigus foliaceus  
 PS Claim 1; Page 10-12; 17pp; Japanese.  
 CC This sequence represents a fused protein recognised by pemphigus  
 CC foliaceus patient autoantibody which comprises the constant region  
 CC of IgG linked to the extracellular region of pemphigus foliaceus  
 CC antigen protein through the hinge portion. Pemphigus foliaceus is  
 CC a chronic, generalised, vesicular and scaling skin eruption similar  
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion  
 CC protein is useful to treat pemphigus foliaceus. The antigen is  
 CC especially administered through an adsorbent upon which the fusion  
 CC protein is immobilised via a carrier. The fusion protein is also  
 CC useful for detecting pemphigus foliaceus antibodies which is useful  
 CC in immunodiagnosis. The fusion protein has little or no side effects.  
 SO Sequence 778 AA;

Query Match 70.5%; Score 55; DB 1; Length 778;  
 Best Local Similarity 60.0%; Pred. No. 0.085;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGFVVDKNTGDINI 15  
 :|||:|  
 DB 97 YGFVVDKNTGELNV 111

RESULT 9  
 R93963  
 ID R93963 standard; peptide; 17 AA.  
 AC R93963;  
 DT 05-JUL-1996 (first entry)  
 DE Peptide combining with anti-interdermal cellular antibody.  
 KW anti-interdermal cellular antibody; autoantibody; adsorbent.  
 OS Synthetic.  
 PN J07309893-A.  
 PD 28-NOV-1995.  
 PF 18-MAY-1994; 129556.  
 PR 18-MAY-1994; JP-129556.  
 PA (KURS ) KURARAY CO LTD.  
 DR WPI; 96-045392/05.  
 PT Anti-Interdermal cellular antibody-combining peptide - which can  
 be immobilised on column to form adsorbent useful for treating  
 diseases related to the antibody.  
 PS Example 3; Page 5; 7pp; Japanese.  
 CC New peptides are disclosed which contain at least 5 contiguous amino  
 CC acids from the sequence Val Gly Ile Asp Gln Pro Phe Gly Ile Phe  
 CC Val Asp Lys Asp Thr Gly Asp Ile Asp Ile Thr Ala Ile Val Asp Arg  
 CC Glu Glu (see R93960), the peptide not containing more than 50 residues.  
 CC The peptide combines with anti-interdermal cellular antibody. It can

CC be immobilised on a carrier to prepare an absorbent useful for the  
 CC treatment of diseases related to anti-interleukin cellular antibody.  
 CC The present sequence is a specific example of the new peptides.  
 SQ Sequence 17 AA;

Query Match 61.5%; Score 48; DB 1; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 0.019;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 DKNTGDINI 15  
 |||||  
 DB 1 DKNTGDINI 9

RESULT 10  
 W13136  
 ID W13136 standard; Protein: 713 AA.

AC W13136;  
 DT 14-MAY-1997 (first entry)  
 DE Full length human cadherin-13.  
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
 KW brain; human; antibody; purification; determination;  
 KW tissue expression; binding antagonist; calcium ion.  
 OS Homo sapiens.  
 PN USS59725-A.  
 PD 28-JAN-1997.  
 PF 17-APR-1992; 872643.  
 PR 17-APR-1992; US-872643.  
 PR 19-APR-1993; US-049460.  
 PR 26-JAN-1994; US-188228.  
 PA (DOHE-) DOHENY EYE INST.  
 PI Suzuki S;  
 DR WPI: 97-108328/10.  
 DR N-PSDB; T61927.  
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
 etc.

PS Example 2: Columns 107-112; 59pp; English.  
 CC The present sequence is full length human cadherin-13, which  
 CC is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA  
 CC was isolated from a foetal brain cDNA library, using probes based  
 CC on homologous rat cadherin cDNA.  
 CC Antibodies or fragments that specifically bind the human cadherin  
 CC can be used to purify the cadherin, determine its tissue expression  
 CC and antagonise its ligand/antigen binding activities.  
 SQ Sequence 713 AA;

Query Match 53.8%; Score 42; DB 1; Length 713;  
 Best Local Similarity 42.9%; Pred. No. 15;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 GIVVDKNTGDINI 15  
 |||:::||||:  
 DB 186 GIVVDKNTGDINI 199

RESULT 11  
 W25638  
 ID W25638 standard; Protein: 713 AA.

AC W25638;  
 DT 03-NOV-1997 (first entry)  
 DE Human cadherin-13.  
 KW Human; cadherin; rat; calcium-dependent cell adhesion protein;  
 KW superfamily; cytoskeleton; eatenin; cancer.  
 OS Homo sapiens.  
 PN Key Location/Qualifiers  
 FT Misc\_difference 703  
 FT /note- "Encoded by CTG"

PN US5646250-A.  
 PD 08-JUL-1997.  
 PF 17-APR-1992; 872643.  
 PR 19-APR-1993; US-049460.

PR 17-APR-1992; US-872643.  
 PR 01-NOV-1994; US-332638.  
 PA (DOHE-) DOHENY EYE INST.  
 PI Suzuki S;  
 DR WPI: 97-362997/73.  
 DR N-PSDB; T85405.

PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion  
 PS Claim 1: Column 99-102; 56pp; English.  
 CC This sequence represents human cadherin-13. The invention specifically  
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat  
 CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell  
 CC adhesion proteins. They are glycosylated integral membrane proteins  
 CC that have an N-terminal extracellular domain that determines binding  
 CC specificity, a hydrophobic membrane spanning region and a C-terminal  
 CC cytoplasmic domain, which is highly conserved among members of the  
 CC superfamily. The C-terminal domain interacts with the cytoskeleton  
 CC through eatenins and other cytoskeleton-associated proteins. The  
 CC novel cadherin proteins may be used in the analysis of the role of  
 CC cadherins in various cancers. Sequence analysis of the cadherin  
 CC proteins also allows investigation of the structure and function of  
 CC cadherin. The cadherin proteins may also be used to modulate the activity  
 CC of cadherin and to determine the tissue specific distribution of cadherin  
 CC proteins. Each subclass of cadherins has a unique tissue distribution  
 CC pattern.  
 SQ Sequence 713 AA;

Query Match 53.8%; Score 42; DB 1; Length 713;  
 Best Local Similarity 42.9%; Pred. No. 15;  
 Matches 6; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 2 GIVVDKNTGDINI 15  
 |||:::||||:  
 DB 186 GIVVDKNTGDINI 199

RESULT 12

W13135  
 ID W13135 standard; Protein: 794 AA.

AC W13135;  
 DT 14-MAY-1997 (first entry)  
 DE Putative human cadherin-12.  
 KW Ca2+ dependent; cell adhesion protein; foetal; cadherin; rat;  
 KW brain; human; antibody; purification; determination;  
 KW tissue expression; binding antagonist; calcium ion; putative.  
 OS Homo sapiens.  
 PN USS59725-A.  
 PD 28-JAN-1997.  
 PF 17-APR-1992; 872643.  
 PR 17-APR-1992; US-872643.  
 PR 19-APR-1993; US-049460.  
 PR 26-JAN-1994; US-188228.  
 PA (DOHE-) DOHENY EYE INST.  
 PI Suzuki S;  
 DR WPI: 97-108328/10.  
 DR N-PSDB; T61926.  
 PT Antibodies to cadherin proteins - useful as cadherin antagonists,  
 etc.  
 PS Example 2: Columns 101-106; 59pp; English.  
 CC The present sequence is a putative human cadherin-12, which  
 CC is a Ca2+ dependent cell adhesion protein. The human cadherin cDNA  
 CC was isolated from a foetal brain cDNA library, using probes based  
 CC on homologous rat cadherin cDNA.  
 CC Antibodies or fragments that specifically bind the human cadherin  
 CC can be used to purify the cadherin, determine its tissue expression  
 CC and antagonise its ligand/antigen binding activities.  
 SQ Sequence 794 AA;

Query Match 52.6%; Score 41; DB 1; Length 794;  
 Best Local Similarity 50.0%; Pred. No. 26;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 IFVVDKNTGDIN 14  
 DB 101 VFIDERTGDH 112

RESULT 13  
 W25637  
 ID W25637 standard; Protein; 794 AA.  
 AC W25637;  
 DT 03-NOV-1997 (first entry)  
 DE Human cadherin-12.  
 KM Human; cadherin; rat; calcium-dependent cell adhesion protein;  
 OS superfamily; cytoskeleton; eatenin; cancer.  
 FH Homo sapiens.  
 FH Key Location/Qualifiers  
 FT MISC\_difference 287  
 FT /note- "Encoded by GCR"  
 FT MISC\_difference 778  
 FT /note- "Encoded by TGC"  
 FT  
 PN US5646250-A.  
 PD 08-JUL-1997.  
 PF 17-APR-1992; 872643.  
 PR 19-APR-1993; US-049460.  
 PR 17-APR-1992; US-872643.  
 PR 01-NOV-1994; US-332638.  
 PA (DOHE-) DOHENY EYE INST.  
 PI Suzuki S.  
 DR WPI: 97-362997/33.  
 DR N-PSDB: T85404.  
 PT Human and rat cadherin polypeptide(s) - mediate cell-cell adhesion  
 PS Claim 1; column 95-100; 56pp; English.  
 CC This sequence represents human cadherin-12. The invention specifically  
 CC provides details of human cadherin-5, -8, -11, -12 and -13, and rat  
 CC cadherin-8, -5, -11 and -13. Cadherins are calcium-dependent cell  
 CC adhesion proteins. They are glycosylated integral membrane proteins  
 CC that have an N-terminal extracellular domain that determines binding  
 CC specificity, a hydrophobic membrane spanning region and a C-terminal  
 CC cytoplasmic domain, which is highly conserved among members of the  
 CC superfamily. The C-terminal domain interacts with the cytoskeleton  
 CC through eatenins and other cytoskeleton-associated proteins. The  
 CC novel cadherin proteins may be used in the analysis of the role of  
 CC cadherins in various cancers. Sequence analysis of the cadherin  
 CC proteins also allows investigation of the structure and function of  
 CC cadherin. The cadherin proteins may be isolated by using anti-cadherin  
 CC antibodies. These antibodies may also be used to modulate the activity  
 CC of cadherin and to determine the tissue specific distribution of cadherin  
 CC proteins. Each subclass of cadherins has a unique tissue distribution  
 CC pattern.  
 SO Sequence 794 AA;

Query Match 52.6%; Score 41; DB 1; Length 794;  
 Best Local Similarity 50.0%; Pred. No. 26;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 IFVVDKNTGDIN 14  
 DB 101 VFIDERTGDH 112

RESULT 14  
 R28822  
 ID R28822 standard; Protein; 1091 AA.  
 AC R28822;  
 DT 23-MAR-1993 (first entry)  
 DE Alpha 6B integrin subunit.  
 KM Human; alpha 6A; alpha 6B; integrin; cell surface receptor; adhesion;  
 KM extracellular matrix; cytoskeleton; heterodimer; laminin receptor;  
 OS immunoprecipitation; JAR; choriocarcinoma.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT domain 1012..1037

FT peptide /label- Transmembrane\_domain  
 FT 1..23  
 FT /label- Signal\_peptide  
 FT protein 24..1091  
 FT /label- Mature\_protein  
 FT modified\_site 223  
 FT /label- Glycosylation\_site  
 FT modified\_site 284  
 FT /label- Glycosylation\_site  
 FT modified\_site 370  
 FT /label- Glycosylation\_site  
 FT modified\_site 513  
 FT /label- Glycosylation\_site  
 FT modified\_site 731  
 FT /label- Glycosylation\_site  
 FT modified\_site 748  
 FT /label- Glycosylation\_site  
 FT modified\_site 891  
 FT /label- Glycosylation\_site  
 FT modified\_site 927  
 FT /label- Glycosylation\_site  
 FT modified\_site 958  
 FT /label- Glycosylation\_site  
 FT binding\_site 230..238  
 FT /note- "Putative cation binding domain"  
 FT binding\_site 324..332  
 FT /note- "Putative cation binding domain"  
 FT binding\_site 386..394  
 FT /note- "Putative cation binding domain"  
 FT binding\_site 441..449  
 FT /note- "Putative cation binding domain"  
 FT domain 1040..1044  
 FT /label- Cytoplasmic\_domain  
 FT /note- "Conserved in virtually all integrins"  
 FT misc\_difference 1044..1045  
 FT /note- "Position of deletion of alpha 6A"

PN W09219647-A.  
 PD 12-NOV-1992.  
 PF 27-APR-1992; U03527.  
 PR 03-MAY-1991; US-695564.  
 PA (SCRI) SCRIPPS RES INST.  
 PI Quaranta V, Tamura RN;  
 DR WPI: 92-398799/48.  
 DR N-PSDB: Q31189.  
 PT Integrin alpha sub-unit cytoplasmic domain polypeptide(s) - used  
 PT for prodn. of antibodies and in detection of integrin sub-units  
 PT in body samples  
 PS Disclosure: Page 78-82; 115pp; English.  
 CC The sequences given in R28821-22 are the human alpha 6A and 6B  
 CC integrin subunits. Integrins are a family of cell surface receptors  
 CC which serve cellular adhesion functions. These receptors form a link  
 CC between the extracellular matrix and the cytoskeleton through their  
 CC binding to various extracellular components. Each integrin receptor  
 CC is a heterodimer comprised of an alpha and a beta subunit. Each alpha  
 CC subunit tends to associate with only one type of beta subunit but  
 CC there are several exceptions to this rule. These integrins correspond  
 CC to the laminin receptor. The cytoplasmic domain of the 6A and 6B  
 CC integrins differs from previously isolated alpha 6 integrins. The  
 CC human alpha 6B was isolated from human choriocarcinoma cell line JAR  
 CC by immunoprecipitation studies.  
 SO Sequence 1091 AA;

Query Match 50.0%; Score 39; DB 1; Length 1091;  
 Best Local Similarity 71.4%; Pred. No. 83;  
 Matches 10; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

QY 1 FGIIVDKNTGDIN 14  
 DB 377 FGIAV-KNIGBAN 388

RESULT 15



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:37 ; Search time 65.16 Seconds  
(without alignments)  
10.858 Million cell updates/sec

Title: US-08-991-628-2  
Perfect score: 78  
Sequence: 1 FGIFVVDKNTGDIINI 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR-62.\*

Word size: 0

Number of hits that pass the threshold: 142080

1: p1r1.\*  
2: p1r2.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	78	100.0	999	1	IJHUG3	desmoglein 3 precu
2	56	71.8	1117	2	S38673	desmoglein 2 - hum
3	55	70.5	1049	1	IJHUG1	desmoglein 1 precu
4	55	70.5	1043	1	IJBGG1	desmoglein 1 precu
5	51	65.4	785	2	I50180	cadherin-7 - chick
6	45	57.7	761	1	IJBODF	desmocollin 1a - b
7	45	57.7	839	1	IJBODF	desmocollin 1b pre
8	45	57.7	770	2	B48910	desmocollin 1b pre
9	45	57.7	824	2	A48910	desmocollin 1a pre
10	45	57.7	840	2	I37281	Desca precursor -
11	45	57.7	894	2	I37282	Desca precursor -
12	43	55.1	790	2	I37016	cadherin-6 - human
13	43	55.1	789	2	I52701	K-cadherin - rat
14	43	55.1	790	2	I50178	cadherin-68 - Chic
15	43	55.1	790	2	G02678	cadherin-14 - huma
16	43	55.1	214	2	S76379	hypothetical prote
17	43	55.1	566	2	S64005	hypothetical prote
18	42	53.8	713	2	B38992	cadherin 13 precu
19	42	53.8	1072	2	T00041	BH-protocadherin P
20	42	53.8	1200	2	T00042	BH-protocadherin P
21	42	53.8	1069	2	T00040	BH-protocadherin P
22	41	52.6	794	2	I59372	cadherin 12 - huma
23	40	51.3	266	2	A46610	multicatalytic end
24	40	51.3	1626	2	A39242	DNA topoisomerase
25	40	51.3	1612	2	S59869	N utilization subs
26	40	51.3	344	2	H72213	Aldehyde ferredoxi
27	40	51.3	598	2	D69292	Bombyx mori recept
28	39	50.0	1715	2	JE0128	F-cadherin - Afric
29	39	50.0	790	2	I51638	nonheme ferritin h
30	39	50.0	165	2	B64121	El protein - human
31	39	50.0	646	2	S36586	probable serine/th
32	39	50.0	1358	2	S33653	integrin alpha-6 c
33	39	50.0	1073	2	B36429	integrin alpha-6 c
34	39	50.0	1091	2	A41543	integrin alpha-6 c
35	39	50.0	73	2	I51528	integrin alpha 6 s

36	39	50.0	1748	2	S63127	probable membrane
37	38.5	49.4	798	2	S62791	Probable lipoprote
38	38	48.7	881	1	JDYLS	DNA-directed DNA p
39	38	48.7	887	1	ITCHCL	E-cadherin precurs
40	38	48.7	657	1	W1W18	El protein - human
41	38	48.7	456	2	S04079	nitrogenase (EC 1.
42	38	48.7	273	2	B64082	outer membrane pro
43	38	48.7	423	2	T06774	cell division prot
44	38	48.7	828	2	S74450	ferrichrome-iron r
45	38	48.7	626	2	C70104	DNA topoisomerase

## ALIGNMENTS

RESULT 1

IJHUG3  
desmoglein 3 precursor - human  
N:Alternate names: pemphigus vulgaris antigen  
C:Species: Homo sapiens (man)  
C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
R:Accession: A41088  
R:Magal, M.; Klaus-Kovtun, V.; Stanley, J.R.  
Cell 67, 869-877, 1991  
A:Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a  
Cell 67, 869-877, 1991  
A:Reference number: A41088; MID:92069753  
A:Accession: A41088  
A:Molecule type: mRNA  
A:Residues: 1-999 <AMA>  
A:Cross-references: GB:M76482; NID:9190751; PID:AAA60230.1; PID:9190752  
C:Genetics:  
A:Gene: GDB:DSG3  
A:Cross-references: GDB:134030; OMIM:169615  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-999/Domain: desmoglein homolog #status predicted <MAT>  
F:50-615/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-267/Domain: cadherin repeat homology <CR2>  
F:270-383/Domain: cadherin repeat homology <CR3>  
F:390-495/Domain: cadherin repeat homology <CR4>  
F:496-598/Domain: cadherin repeat homology <CR5>  
F:616-939/Domain: transmembrane #status predicted <TM>  
F:640-939/Domain: intracellular #status predicted <INT>  
F:910-938/Domain: desmoglein repeat <DG1>  
F:937-966/Domain: desmoglein repeat <DG2>  
F:110,180,545/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 100.0%; Score 78; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 3.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDIINI 15  
DB 97 FGIFVVDKNTGDIINI 111

RESULT 2  
S38673  
desmoglein 2 - human  
N:Alternate names: desmoglein HDG  
C:Species: Homo sapiens (man)  
C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999  
C:Accession: S38673; B38872  
R:Zimbleman, R.  
Submitted to the EMBL-Data Library, September 1993  
A:Reference number: S38673  
A:Accession: S38673  
A:Status: preliminary



A:Molecule type: mRNA  
A:Residues: 1-1117 <ZIM>  
A:Cross-references: EMBL:Z26317; NID:9416177; PIDN:CAA81226.1; PID:9416178  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Eur. J. Cell Biol. 55, 200-208, 1991  
A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide  
A:Reference number: A38872; MUID:92037656  
A:Accession: B38872  
A:Molecule type: mRNA  
A:Residues: 777-1117 <ROC>  
A:Cross-references: GB:S64273  
C:Genetics:  
A:Gene: GDB:DSG2  
A:Cross-references: GDB:128808; OMIM:125671  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein  
F:51-158/Domain: cadherin repeat homology <CR1>  
F:161-271/Domain: cadherin repeat homology <CR2>

Query Match 71.8%; Score 56; DB 2; Length 1117;  
Best Local Similarity 60.0%; Pred. No. 0.19;  
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDINI 15  
|||||:|:|:|:|:  
DB 98 FGIFVFNKDTGELNV 112

RESULT 3  
ITBHG1  
desmoglein 1 precursor - human  
N:Alternate names: desmosomal glycoprotein I  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence; revision 30-Jun-1993 #text; change 22-Jun-1999  
C:Accession: S16906; A39706; A61254; A61279; S16158  
R:Buxton, R.S.  
Submitted to the EMBL Data Library, November 1990  
A:Reference number: S16906  
A:Accession: S16906  
A:Molecule type: mRNA  
A:Residues: 1-1049 <BOX>  
A:Cross-references: EMBL:X56654; NID:930505; PIDN:CAA39976.1; PID:930506  
R:Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Ataliotis, P.; Poynter, D.; Arneemann, J.; F  
Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991  
A:Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,  
A:Reference number: A39706; MUID:91271279  
A:Accession: A39706  
A:Molecule type: mRNA  
A:Residues: 24-1049 <WHE>  
A:Cross-references: GB:X56654  
R:Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arneemann, J.; Rees, D.A.; King, I.A.; Magee  
Biochem. Soc. Trans. 19, 1060-1064, 1991  
A:Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily  
A:Reference number: A61279; MUID:92175187  
A:Accession: A61279  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-55 <WH3>  
C:Genetics:  
A:Gene: GDB:DSG1  
A:Cross-references: GDB:126563; OMIM:125670  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro

F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-1048/Product: desmoglein #status predicted <MAT>  
F:50-548/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-269/Domain: cadherin repeat homology <CR2>  
F:272-385/Domain: cadherin repeat homology <CR3>  
F:392-493/Domain: cadherin repeat homology <CR4>  
F:509-530/Region: serine/threonine-rich  
F:549-569/Domain: transmembrane #status predicted <TM>  
F:572-1049/Domain: intracellular #status predicted <INT>  
F:840-869/Domain: desmoglein repeat <DG1>  
F:870-899/Domain: desmoglein repeat <DG2>  
F:900-927/Domain: desmoglein repeat <DG3>  
F:928-956/Domain: desmoglein repeat <DG4>  
F:969-1019/Region: glycine/serine-rich  
F:110-180/Binding site: carboxydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 55; DB 1; Length 1049;  
Best Local Similarity 60.0%; Pred. No. 0.26;  
Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FGIFVVDKNTGDINI 15  
|||||:|:|:|:|:  
DB 97 YGIFVFNKDTGELNI 111

RESULT 4  
ITBHG1  
desmoglein 1 precursor - bovine  
N:Alternate names: desmoglein BDGM  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 30-Jun-1993 #sequence; revision 30-Jun-1993 #text; change 22-Jun-1999  
C:Accession: S14603; A38872; A37785; S38721; A48173; S24412  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Submitted to the EMBL Data Library, March 1991  
A:Description: Complete sequence of the desmoglein precursor and evidence for the exl  
A:Reference number: S14603  
A:Accession: S14603  
A:Molecule type: mRNA  
A:Residues: 1-1043 <KOC>  
A:Cross-references: EMBL:X58466; NID:9306; PIDN:CAA41380.1; PID:9307  
R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
Eur. J. Cell Biol. 55, 200-208, 1991  
A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypept  
A:Reference number: A38872; MUID:92037656  
A:Accession: A38872  
A:Molecule type: mRNA  
A:Residues: 1-87;968-1043 <KO2>  
A:Cross-references: GB:S64268; GB:S64270  
R:Goodwin, D.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowlin, P.  
Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990  
A:Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion  
A:Reference number: A37785; MUID:91097553  
A:Accession: A37785  
A:Molecule type: mRNA  
A:Residues: 44-123; 125-493 <GOO>  
A:Cross-references: GB:M8165; NID:9162966; PIDN:AAA62709.1; PID:9552318  
R:Zimbelmann, R.  
Submitted to the EMBL Data Library, February 1991  
A:Reference number: S38721  
A:Accession: S38721  
A:Molecule type: mRNA  
A:Residues: 44-1043 <ZIM>  
A:Cross-references: EMBL:X57784; NID:9436061; PIDN:CAA40930.1; PID:9436062  
R:Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.  
Eur. J. Cell Biol. 53, 1-12, 1990  
A:Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m  
A:Reference number: A48173; MUID:91166965  
A:Accession: A48173  
A:Molecule type: mRNA  
A:Residues: 44-1001; AQPSPAT' <KO3>

A:Cross-references: GB:X57784  
 A>Note: this sequence has been revised in references A38872 and S38721  
 C:Genetics:  
 A:Gene: DSG1  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane protein  
 F:1-73/Domain: signal sequence #status predicted <SIG>  
 F:24-49/Domain: propeptide #status predicted <PRO>  
 F:50-1003/Product: desmoglein #status predicted <MAT>  
 F:50-548/Domain: extracellular #status predicted <EXT>  
 F:52-157/Domain: cadherin repeat homology <CR1>  
 F:160-269/Domain: cadherin repeat homology <CR2>  
 F:272-385/Domain: cadherin repeat homology <CR3>  
 F:392-491/Domain: cadherin repeat homology <CR4>  
 F:548-574/Domain: transmembrane #status predicted <TM>  
 F:574-1043/Domain: intracellular #status predicted <INT>  
 F:846-875/Domain: desmoglein repeat <DG1>  
 F:876-905/Domain: desmoglein repeat <DG2>  
 F:906-933/Domain: desmoglein repeat <DG3>  
 F:934-962/Domain: desmoglein repeat <DG4>  
 F:963-1012/Region: glycine/serine-rich  
 F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5% Score 55; DB 1; Length 1043;  
 Best Local Similarity 60.0% Pred. No. 0.26; Indels 0; Gaps 0;  
 Matches 9; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 1 FGIFVVKNTGDI 15  
 DB 97 YGIFVVKNTGDI 111

RESULT 5  
 150180  
 C:Species: Gallus gallus (chicken)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 20-Aug-1999  
 C:Accession: 150180  
 R:Nakagawa, S.; Takeuchi, M.  
 Development 121, 1321-1332, 1995  
 A>Title: Neural crest cell-cell adhesion controlled by sequential and subpopulation-specific  
 A:Reference number: 150178; MUID:95509115  
 A:Accession: 150180  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-785 <NAK>  
 A:Cross-references: GB:DA2150; NID:9868000; PIDN:BA07721.1; PID:9868001  
 C:Superfamily: cadherin; cadherin repeat homology

Query Match 65.4% Score 51; DB 2; Length 785;  
 Best Local Similarity 66.7% Pred. No. 0.91; Indels 0; Gaps 0;  
 Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 IFVVDKNTGDI 14  
 DB 94 IFVVDKNTGDI 105

RESULT 6  
 10506  
 N:Alternate names: desmocollin BDCM; desmosomal glycoprotein 2  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
 C:Accession: A43838; A38872; A38456; A60714; S14542  
 R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbermann, R.; Schmelz, M.; Franke, W.W.  
 Differentiation 47, 29-36, 1991  
 A>Title: Amino acid sequence of bovine muzzle epithelial desmocollin derived from cloned  
 A:Reference number: A43838; MUID:92008912  
 A:Accession: A43838

A:Molecule type: mRNA  
 A:Residues: 1-761 <KOC>  
 A:Cross-references: GB:X58029; GB:S57985; NID:9453583; PIDN:CAA4088.1; PID:9453584  
 A:Accession: B43838  
 A:Molecule type: Protein  
 A:Residues: 1-32;65-76;148-159;164-176;190-205;208-219;238-256;361-375;377-388;478-48  
 A:Experimental source: muzzle epithelium  
 A>Note: sequence extracted from NCBI database  
 R:Collins, J.E.; Legan, P.K.; Kenny, T.P.; Macgarvie, J.; Holton, J.L.; Garrod, D.R.  
 J. Cell Biol. 113, 381-391, 1991  
 A>Title: Cloning and sequence analysis of desmosomal glycoproteins 2 and 3 (desmocoll  
 A:Reference number: A38456; MUID:91185414  
 A:Accession: A38456  
 A:Molecule type: mRNA  
 A:Residues: 606-761 <COL>  
 A:Cross-references: EMBL:X56967; NID:9310; PIDN:CAA40287.1; PID:9311  
 R:Holton, J.L.; Kenny, T.P.; Legan, P.K.; Collins, J.E.; Keen, J.N.; Sharma, R.; Garr  
 J. Cell Sci. 97, 239-246, 1990  
 A>Title: Desmosomal glycoproteins 2 and 3 (desmocollins) show N-terminal similarity t  
 A:Reference number: A60714; MUID:9115997  
 A:Accession: A60714  
 A:Molecule type: Protein  
 A:Residues: 1-6, 'A', '8-9', 'R', '11-17', 'RCE', '21-23 <HOL>  
 A:Experimental source: nasal epidermis  
 C:Genetics:  
 A:Gene: DSC1  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycop  
 F:1-761/Product: desmocollin 1a #status experimental <MAT>  
 F:1-561/Domain: extracellular #status predicted <EXT>  
 F:3-108/Domain: cadherin repeat homology <CR1>  
 F:111-220/Domain: cadherin repeat homology <CR2>  
 F:223-338/Domain: cadherin repeat homology <CR3>  
 F:339-444/Domain: cadherin repeat homology <CR4>  
 F:445-561/Domain: cadherin repeat homology <CR5>  
 F:562-582/Domain: transmembrane #status predicted <TM>  
 F:583-761/Domain: intracellular #status predicted <INT>  
 F:31.266.413/Binding site: carbohydrate (Asn) (covalent) (by protein kinase C) #status  
 F:584.588.678/Binding site: phosphate (Thr) (covalent) (by protein kinase II) #status predi  
 F:605/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predi  
 F:671/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predi  
 F:681/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:682/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predi

Query Match 57.7% Score 45; DB 1; Length 761;  
 Best Local Similarity 53.8% Pred. No. 8.9; Indels 0; Gaps 0;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 FGIFVVKNTGDI 13  
 DB 48 FVDFVVKNTGDI 60

RESULT 7  
 10506  
 N:Alternate names: desmosomal glycoprotein 3  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
 C:Accession: B38456; A39377; S14567  
 R:Collins, J.E.; Legan, P.K.; Kenny, T.P.; Macgarvie, J.; Holton, J.L.; Garrod, D.R.  
 J. Cell Biol. 113, 381-391, 1991  
 A>Title: Cloning and sequence analysis of desmosomal glycoproteins 2 and 3 (desmocoll  
 A:Reference number: A38456; MUID:91185414  
 A:Accession: B38456  
 A:Molecule type: mRNA  
 A:Residues: 1-839 <COL>  
 A:Cross-references: GB:X56966; NID:9315; PIDN:CAA40286.1; PID:9316  
 R:Mechanic, S.; Raynor, K.; Hill, J.E.; Cowin, P.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 4476-4480, 1991  
 A>Title: Desmocollins form a distinct subset of the cadherin family of cell adhesion  
 A:Reference number: A39377; MUID:91239591

A:Accession: A39377  
A:Molecule type: mRNA  
A:Residues: 1-484, 'A', 486-839 <MEC>  
A:Cross-references: GB:M67489; GB:M61750; NID:g162970; PIDN:AAA30492.1; PID:g162971  
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed by sequencing.  
C:Gene: DSC1  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: alternative splicing; calcium binding; cell adhesion; duplication; glycoprotein  
E:1-29/Domain: signal sequence #status predicted <STG>  
E:30-132/Domain: propeptide #status predicted <PRO>  
E:133-839/Product: desmocollin 1b #status experimental <MAT>  
F:133-682/Domain: extracellular #status predicted <EXT>  
F:135-240/Domain: cadherin repeat homology <CR1>  
F:243-352/Domain: cadherin repeat homology <CR2>  
F:355-470/Domain: cadherin repeat homology <CR3>  
F:471-576/Domain: cadherin repeat homology <CR4>  
F:577-682/Domain: cadherin repeat homology <CR5>  
F:694-714/Domain: transmembrane #status predicted <TM>  
F:718-839/Domain: intracellular #status predicted <INT>  
F:163-398,545/Binding site: carboxylate (Asn) (covalent) #status predicted  
F:716,720,810/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:737/Binding site: phosphate (Ser) (covalent) (by protein kinase II) #status predicted  
F:803,830/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted  
F:813/Binding site: phosphate (Tyr) (covalent) #status predicted  
F:814/Binding site: phosphate (Thr) (covalent) (by protein kinase II) #status predicted

Query Match 57.7%; Score 45; DB 1; Length 839;  
Best Local Similarity 53.8%; Pred. No. 9.9;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDI 13  
| : : : : :  
Db 180 FNLFIYKDTGDI 192

RESULT 8  
B48910  
desmocollin 1b precursor - human (fragment)  
C:Species: Homo sapiens (man)  
C>Date: 08-Dec-1995 #sequence\_revision 08-Dec-1995 #text\_change 15-Aug-1997  
C:Accession: B48910  
R:King, I.A.; Arneemann, J.; Spurr, N.K.; Buxton, R.S.  
Genomics 18, 185-194, 1993  
A:Title: Cloning of the cDNA (DSC1) coding for human type 1 desmocollin and its assignment to the desmocollin gene.  
A:Reference number: A48910; MUID:94116981  
A:Accession: B48910  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-770 <KIN>  
A:Cross-references: GB:X72929  
C:Gene: DSC1  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: alternative splicing; calcium binding; cell adhesion; glycoprotein; phosphoglycoprotein

Query Match 57.7%; Score 45; DB 2; Length 770;  
Best Local Similarity 53.8%; Pred. No. 9;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDI 13  
| : : : : :  
Db 112 FNLFIYKDTGDI 124

RESULT 9  
A48910  
desmocollin 1a precursor - human (fragment)

C:Species: Homo sapiens (man)  
C>Date: 08-Dec-1995 #sequence\_revision 08-Dec-1995 #text\_change 15-Aug-1997  
C:Accession: A48910  
R:King, I.A.; Arneemann, J.; Spurr, N.K.; Buxton, R.S.  
Genomics 18, 185-194, 1993  
A:Title: Cloning of the cDNA (DSC1) coding for human type 1 desmocollin and its assignment to the desmocollin gene.  
A:Reference number: A48910; MUID:94116981  
A:Accession: A48910  
A:Status: Preliminary  
A:Molecule type: mRNA  
A:Residues: 1-824 <KIN>  
A:Cross-references: GB:X72925  
C:Gene: DSC1  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: alternative splicing; calcium binding; cell adhesion; glycoprotein; phosphoglycoprotein

Query Match 57.7%; Score 45; DB 2; Length 824;  
Best Local Similarity 53.8%; Pred. No. 9.7;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDI 13  
| : : : : :  
Db 112 FNLFIYKDTGDI 124

RESULT 10  
I37281  
Dsc1 precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 20-Aug-1999  
C:Accession: I37281  
R:Thiels, D.G.; Koch, P.V.; Franke, W.W.  
Int. J. Dev. Biol. 37, 101-110, 1993  
A:Title: Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified squamous epithelium.  
A:Reference number: I37281; MUID:93283249  
A:Accession: I37281  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-840 <RES>  
A:Cross-references: EMBL:Z34522; NID:g505536; PIDN:CAA84279.1; PID:g505537  
C:Superfamily: cadherin; cadherin repeat homology

Query Match 57.7%; Score 45; DB 2; Length 840;  
Best Local Similarity 53.8%; Pred. No. 9.9;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 FGIFVVDKNTGDI 13  
| : : : : :  
Db 182 FNLFIYKDTGDI 194

RESULT 11  
I37282  
Dsc1b precursor - human  
C:Species: Homo sapiens (man)  
C>Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 20-Aug-1999  
C:Accession: I37282  
R:Thiels, D.G.; Koch, P.V.; Franke, W.W.  
Int. J. Dev. Biol. 37, 101-110, 1993  
A:Title: Differential synthesis of type 1 and type 2 desmocollin mRNAs in human stratified squamous epithelium.  
A:Reference number: I37281; MUID:93283249  
A:Accession: I37282  
A:Status: Preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-894 <RES>  
A:Cross-references: EMBL:Z34522; NID:g505536; PIDN:CAA84279.1; PID:g505538  
C:Gene: DSC1

A:introns: 829/2  
C:Superfamily: cadherin; cadherin repeat homology

Query Match 57.7%; Score 45; DB 2; Length 894;  
Best Local Similarity 53.8%; Pred. No. 11;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 IFVVDKNTGDI 13  
|:|:|:|:|:|:|  
Db 182 FNLFLYKEDTGD 194

RESULT 12  
137016

cadherin-6 - human

C:Species: Homo sapiens (man)  
C>Date: 06-Sep-1996 #sequence\_revision 06-Sep-1996 #text\_change 20-Aug-1999

C:Accession: I37016

R:Shimoyama, Y.; Gotoh, M.; Terasaki, T.; Kitajima, M.; Hirohashi, S.

Cancer Res. 55, 2206-2211, 1995

A:Title: Isolation and sequence analysis of human cadherin-6 complementary DNA for the

A:Reference number: I37016; MUID:95262134

A:Accession: I37016

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-790 <RES>  
A:Cross-references: GB:D31784; NID:974184; PIDN:BA06562.1; PID:974185

C:Genetics:

A:Gene: GDB:CDH6

A:Cross-references: GDB:5822908

C:Superfamily: cadherin; cadherin repeat homology

Query Match 55.1%; Score 43; DB 2; Length 790;  
Best Local Similarity 54.5%; Pred. No. 20;  
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 13  
|:|:|:|:|:|:|  
Db 100 LFIINNTGDI 110

RESULT 13  
152701

K-cadherin - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 20-Aug-1999

C:Accession: I52701

R:Xiang, Y.

Cancer Res. 54, 3034-3041, 1993

A:Title: Isolation of complementary DNA encoding K-cadherin, a novel rat cadherin prefer

A:Reference number: I52701; MUID:94243827

A:Accession: I52701

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-789 <RES>  
A:Cross-references: GB:D25290; NID:9435460; PIDN:BA04975.1; PID:9435461

C:Genetics:

A:Gene: KCAD

C:Superfamily: cadherin; cadherin repeat homology

Query Match 55.1%; Score 43; DB 2; Length 789;  
Best Local Similarity 54.5%; Pred. No. 20;  
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 13  
|:|:|:|:|:|:|  
Db 100 LFIINNTGDI 110

RESULT 14

150178

cadherin-6B - chicken

C:Species: Gallus gallus (chicken)

C>Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 20-Aug-1999

C:Accession: I50178

R:Nakagawa, S.; Takeichi, M.

Development 121, 1321-1332, 1995

A:Title: Neural crest cell-cell adhesion controlled by sequential and subpopulation-s

A:Reference number: I50178; MUID:95309115

A:Accession: I50178

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-790 <NAK>

A:Cross-references: GB:D42149; NID:9867998; PIDN:BA07720.1; PID:9867999

C:Superfamily: cadherin; cadherin repeat homology

Query Match 55.1%; Score 43; DB 2; Length 790;  
Best Local Similarity 54.5%; Pred. No. 20;  
Matches 6; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 13  
|:|:|:|:|:|:|  
Db 100 LFIINNTGDI 110

RESULT 15  
G02678

cadherin-14 - human

C:Species: Homo sapiens (man)

C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 20-Aug-1999

C:Accession: G02678

R:Shimoyama, Y.; Gotoh, M.; Hirohashi, S.

submitted to the EMBL Data Library, May 1996

A:Reference number: H01584

A:Accession: G02678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-790 <SHI>  
A:Cross-references: EMBL:U59325; NID:91389852; PIDN:AA02933.1; PID:91389853

C:Superfamily: cadherin; cadherin repeat homology

Query Match 55.1%; Score 43; DB 2; Length 790;  
Best Local Similarity 58.3%; Pred. No. 20;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 3 IFVVDKNTGDI 14  
|:|:|:|:|:|:|  
Db 100 IFIIDDITGDH 111

Search completed: January 7, 2000, 08:52:39  
Job time: 2269 sec

Sat Jan 15 11:44:57 2000

us-08-991-628-2.rpr

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:40:34 ; Search time 51.33 Seconds  
(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-3

Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36.\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	71	100.0	999	1 R30742	Human pemphigus vu
2	71	100.0	614	1 W07908	pemphigus vulgaris
3	71	100.0	15	1 W04843	Self epitope of de
4	71	100.0	15	1 W64815	Desmoglein-3 190-2
5	71	100.0	15	1 W78814	Desmoglein 3 prote
6	62	87.3	778	1 W15489	Pemphigus foliaceu
7	55	77.5	263	1 W13009	Segment of desmoso
8	55	77.5	263	1 W13010	Segment of desmoso
9	39	54.9	878	1 R85487	Human E-cadherin p
10	39	54.9	878	1 R55060	Sequence of human
11	34	47.9	4472	1 R97246	Virulence gene clu
12	34	47.9	2233	1 W48711	HPV-3 US isolate
13	34	47.9	2233	1 W48712	HPV-3 FRh1 cp45 v
14	34	47.9	2233	1 W48713	HPV-3 Vero cp45 v
15	34	47.9	456	1 W82841	Human cerebral pro
16	34	47.9	330	1 W89745	Staphylococcus aur
17	34	46.5	2408	1 R24307	Translocation of ORF
18	33	46.5	479	1 W23241	P. membranefaciens
19	33	46.5	81	1 W75861	Human secretory pr
20	33	46.5	438	1 W95500	B. subtilis yabE r
21	32	45.1	697	1 R58591	Hepatitis C virus
22	32	45.1	730	1 R82881	Lupin exo-(1-4)bet
23	32	45.1	532	1 R91035	Recombinant V8 pro
24	32	45.1	611	1 W02157	Periplasmic Beta-N
25	32	45.1	916	1 W13129	Full length human
26	32	45.1	205	1 W20360	H. pylori secreted
27	32	45.1	217	1 W20960	H. pylori secreted
28	32	45.1	537	1 W22320	Protein encoded by
29	32	45.1	645	1 W20054	S. aureus threonyl
30	32	45.1	532	1 W22219	Protein encoded by
31	32	45.1	916	1 W25658	Human cadherin-4.
32	32	45.1	481	1 W34554	MLTR glycosidase
33	32	45.1	419	1 W30917	DNA gyrase subunit
34	32	45.1	371	1 W41513	N. meningitidis al
35	32	45.1	371	1 W41514	N. gonorrhoeae alph
36	32	45.1	151	1 W52130	Putative insecticid
37	32	45.1	277	1 W51011	Human liver carbon
38	32	45.1	276	1 W37921	Bos taurus Dnase I
39	32	45.1	26	1 W60192	Bacteriophage spoI

## ALIGNMENTS

40	32	45.1	729	1	W60724	Human TK2 protein
41	32	45.1	398	1	W30523	Kurthia sp. KAPA s
42	32	45.1	645	1	W61633	Staphylococcus aur
43	32	45.1	481	1	W45858	Desulfurococcus HI
44	32	45.1	729	1	W79273	Human T2K (TRAF2-a
45	32	45.1	611	1	W85599	Hexosaminidase enz

RESULT 1  
ID R30742  
AC R30742;  
DE 14-JUN-1993 (first entry)  
DE Human pemphigus vulgaris 130KD antigen.  
KW pemphigus vulgaris; skin disease; autoantibodies;  
KW keratinocyte cell surface antigen; glycoprotein; cell adhesion.  
OS Homo sapiens.  
PN US798918-A.  
PD 15-DEC-1992.  
PE 27-NOV-1991; 798918.  
PR 27-NOV-1991; US-798918.  
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.  
PI Amagai M, Klaus-Kovtun V, Stanley JR;  
DR WPI, 93-067436/08.  
DR N-PSDB; Q35992.  
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for  
PT diagnostic and therapeutic uses  
PS Disclosure; Fig 7; 50PP; English.  
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein  
CC and its encoding DNA may be used in the diagnosis and treatment of  
CC pemphigus vulgaris. It is thought that the antigen may be a cell  
CC adhesion molecule.  
SQ Sequence 999 AA;

Query Match 100.0%; Score 71; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNSKIAFKIVSOEPA 15  
DB 190 LNSKIAFKIVSOEPA 204

RESULT 2  
ID W07908  
AC W07908 standard; protein; 614 AA.  
DE 29-JUN-1997 (first entry)  
DE pemphigus vulgaris antigen protein extracellular region.  
KW Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;  
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;  
KW dermatology.  
OS Homo sapiens.  
PN J08188540-A.  
PD 23-JUL-1996.  
PE 30-JUN-1995; 165632.  
PR 30-JUN-1994; JP-173291.  
PA (NISH/) NISHIKAWA T.  
DR WPI, 96-388562/39.  
PT fused protein recognised by pemphigus vulgaris autoantibody -  
PT useful to treat and diagnose pemphigus vulgaris  
PS Claim 1; Page 7-9; 9PP; Japanese.  
CC W07908 represents the human pemphigus vulgaris (PV) antigen  
CC extracellular region. The PV antigen is produced in patients with  
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare  
CC relapsing disease causing suprabasal, intra-epidermal bullae  
CC (vesicles) of the skin and mucous membranes, which is fatal if  
CC untreated. The PV antigen was fused to a human IgG1 hinge region  
CC and the resulting fusion protein is useful to treat or diagnose

CC pemphigus vulgaris.  
SQ Sequence 614 AA;

Query Match 100.0%; Score 71; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
DB 189 LNSKIAFKIVSOEPA 203

## RESULT 3

W04843  
ID W04843 standard; peptide; 15 AA.  
AC W04843;  
DT 18-FEB-1997 (first entry)  
DE Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerisation; self-epitope; antigen; autoimmune disease;  
KW autoantigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphomannomutase;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW influenza; haemagglutinin; reovirus; sigma protein.  
OS Homo sapiens.  
PN W09627387-A1.  
PD 12-SEP-1996;  
PF 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PA (HARD) HARVARD COLLEGE.  
PI Strominger JL, Muecherfennig KW;  
DR WPI: 96-425218/42.  
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for  
PT identification of other self and non-self antigens implicated in  
PT auto-immune disease  
PS Claim 1; Page 39; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise  
CC either an isolated human non-collagen or non-myosin basic protein  
CC (MBP) polypeptide which is capable of tolerising an individual to an  
CC autoantigen; or an isolated human pathogen polypeptide capable of  
CC tolerising an individual to that polypeptide. In both cases, the  
CC polypeptide (whether self or non-self) includes an amino acid  
CC sequence corresponding to a sequence motif for a MHC class II  
CC protein, such as HLA-DR, which is associated with a human autoimmune  
CC disease and which binds to the polypeptide to activate autoreactive  
CC T-cells in individuals with the autoimmune disease. This peptide is  
CC derived from the human desmoglein 3 protein (amino acids 190-204)  
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
CC derived from the human desmoglein protein are described in W04841-47.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 71; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
DB 1 LNSKIAFKIVSOEPA 15

## RESULT 4

W64815  
ID W64815 standard; peptide; 15 AA.  
AC W64815;  
DT 29-SEP-1998 (first entry)  
DE Desmoglein-3 190-204.  
KW Desmoglein; DG; gene therapy; pemphigus vulgaris; microparticle;  
KW autoantigen; autoimmune disease; MHC.  
OS Homo sapiens.  
PN US5783567-A.

PD 21-JUL-1998.  
PF 22-JAN-1997; 787547.  
PR 22-JAN-1997; US-787547.  
PA (PANG-) PANGAEA PHARM INC.  
PI Curley JM, Hedley ML, Langer RS;  
DR WPI: 98-427077/36.

PT Microparticle encapsulated nucleic acids - for recombinant  
PT expression of proteins e.g. in gene therapy

PS Disclosure; Column 4; 42pp; English.  
CC The patent describes a new preparation of microparticles each  
CC comprising a polymeric matrix and a nucleic acid. The polymeric  
CC matrix consists of one or more synthetic polymers having a solubility  
CC in water of less than 1 mg/l (e.g. poly-lactic-co-glycolic acid);  
CC and at least 90% of the microparticles have a diameter of less than  
CC 100 microns. The microparticles are useful for the delivery of nucleic  
CC acids to phagocytic cells. In one embodiment the microparticles are  
CC less than 20 microns in diameter and the nucleic acid (preferably in  
CC closed circular form) includes an expression control sequence  
CC operatively linked to a coding sequence, where the expression product  
CC of the coding sequence is a polypeptide having a length and a sequence  
CC which permits it to bind to an MHC class I or II molecule. The  
CC expression product is thus an effective stimulator of an immune  
CC response in mammals. The present sequence, an antigenic portion of  
CC desmoglein 3, is an example of an MHC class II peptide which can be  
CC expressed by the nucleic acid. It is associated with pemphigus  
CC vulgaris.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 71; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.4e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
DB 1 LNSKIAFKIVSOEPA 15

## RESULT 5

W78814  
ID W78814 standard; peptide; 15 AA.  
AC W78814;  
DT 17-NOV-1998 (first entry)  
DE Desmoglein 3 protein fragment 190-204.  
KW Microparticle; delivery; polymeric matrix; autoantigen; tumour antigen;  
KW class II associated peptide; pathogen; gene therapy; genetic disease;  
KW infection; downregulation; immune response.  
OS Homo sapiens.  
PN W09831398-A1.  
PD 23-JUL-1998.  
PF 22-JAN-1998; U01499.  
PR 06-JAN-1998; US-003253.  
PR 22-JAN-1997; US-787547.  
PA (PANG-) PANGAEA PHARM INC.  
PI Curley JM, Hedley ML, Langer RS, Lunsford LB;  
DR WPI: 98-427556/36.  
PT New preparations of microparticles - comprising a synthetic polymer  
PT matrix and nucleic acid comprising an expression vector for use in  
PT gene therapy  
PS Disclosure; Page 8; 101pp; English.  
CC A microparticle preparation (MP) has been developed, consisting of  
CC microparticles having a diameter of less than 100 nm. The MP comprises:  
CC (a) a polymeric matrix (PM) consisting of one or more synthetic polymers  
CC having a solubility in water of less than 1 mg/l; and (b) an expression  
CC vector selected from RNA molecules (at least 50% of which are closed  
CC circles) or circular plasmid DNA (at least 50% of which are supercoiled).  
CC Also described is a MP of at most 20 microns in diameter, comprising: (a)  
CC a PM; and (b) a NM comprising an expression control sequence operatively  
CC linked to a coding sequence, where the coding sequence encodes an  
CC expression product selected from: (1) a polypeptide at least 7 amino  
CC acids in length, having a sequence identical to the sequence of: (1) a  
CC fragment of a naturally-occurring mammalian protein; or (1) a fragment

CC of a naturally-occurring protein from an infectious agent which infects  
 CC a mammal; (2) a peptide having a length and sequence which permits it to  
 CC bind to an MHC class I or II molecule; and (3) the polypeptide or the  
 CC peptide linked to a trafficking sequence. W69/63 to W69/65, and W78/793  
 CC to W78/897 are peptide fragments for use in the present invention. The  
 CC MMs are highly effective vehicles for the delivery of polynucleotides  
 CC into phagocytic cells. They can be used for gene therapy, e.g. for  
 CC treating genetic diseases, infections or tumours or for downregulating  
 CC an immune response.  
 SQ Sequence 15 AA;

Query Match 100.0%; Score 71; DB 1; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 14e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNSKIAFKIVSOEPA 15  
 |||||  
 Db 1 LNSKIAFKIVSOEPA 15

RESULT 6  
 W15489  
 ID W15489 standard; Protein; 778 AA.

DT 17-JUN-1997 (first entry)  
 DE Pemphigus foliaceus antigen-IgG constant region fusion protein.  
 KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
 KW extracellular region; antigen; hinge portion; skin;  
 KW dermatitis herpetiformis; fusion protein; detection; ss.  
 OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers  
 FT domain 1..545  
 FT /note="Pemphigus foliaceus antigen protein"  
 FN J09077800-A.

PD 25-MAR-1997.  
 PF 12-SEP-1995; 260899.  
 PR 12-SEP-1995; JP-260899.  
 PA (NISHU) NISHIKAWA T.  
 WP1; 97-241758/22.

DR P-PSDB; T66428.  
 PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked  
 PT through the hinge region used to treat pemphigus foliaceus

PS Claim 1; Page 10-12; 17pp; Japanese.  
 CC This sequence represents a fused protein recognised by pemphigus  
 CC foliaceus patient autoantibody which comprises the constant region  
 CC of IgG linked to the extracellular region of pemphigus foliaceus  
 CC antigen protein through the hinge portion. Pemphigus foliaceus is  
 CC a chronic, generalised, vesicular and scaling skin eruption similar  
 CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion  
 CC protein is useful to treat pemphigus foliaceus. The antigen is  
 CC especially administered through an adsorbent upon which the fusion  
 CC protein is immobilised via a carrier. The fusion protein is also  
 CC useful for detecting pemphigus foliaceus antibodies which is useful  
 CC in immunodiagnosis. The fusion protein has little or no side effects.

CC Sequence 778 AA;  
 SQ

Query Match 87.3%; Score 62; DB 1; Length 778;  
 Best Local Similarity 80.0%; Pred. No. 0.00072;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LNSKIAFKIVSOEPA 15  
 |||||  
 Db 190 LNSKIAFKIRQEPS 204

RESULT 7  
 W13009  
 ID W13009 standard; Protein; 560 AA.  
 AC W13009;  
 DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.

KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;  
 KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.

OS Homo sapiens.  
 PN DE19531033-A1.  
 PD 27-FEB-1997.  
 PF 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.  
 PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;  
 DR WP1; 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on  
 PT surface of epithelial or carcinoma cells, not bound to desmosomes,  
 PT useful for diagnosis and treatment of carcinoma micrometastases  
 PS Claim 7; Page 5; 8pp; German.

CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or  
 CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to  
 CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separate, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to  
 CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or  
 CC carcinomas.  
 SQ Sequence 560 AA;

Query Match 77.5%; Score 55; DB 1; Length 560;  
 Best Local Similarity 73.3%; Pred. No. 0.011;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 LNSKIAFKIVSOEPA 15  
 |||||  
 Db 143 LNSKISYRIVLEPA 157

RESULT 8  
 W13010  
 ID W13010 standard; Protein; 263 AA.  
 AC W13010;

DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.  
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;  
 KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.

OS Homo sapiens.  
 PN DE19531033-A1.  
 PD 27-FEB-1997.  
 PF 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.  
 PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;  
 DR WP1; 97-146518/14.

PT Antibody reactive with part of desmosomal cadherin - exposed on  
 PT surface of epithelial or carcinoma cells, not bound to desmosomes,  
 PT useful for diagnosis and treatment of carcinoma micrometastases  
 PS Claim 9; Page 5; 8pp; German.  
 CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or  
 CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to  
 CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separate, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to  
 CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or



CC Carcinomas.  
SQ Sequence 263 AA:

Query Match 77.5%; Score 55; DB 1; Length 263;  
Best Local Similarity 73.3%; Pred. No. 0.0046;  
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
DB 41 LNSKISIRIVSDEPA 55

RESULT 9  
R55060  
ID R55060 standard; Protein: 878 AA.  
AC R55060.  
DE 08-NOV-1994 (first entry)  
KW Human epithelial-cadherin; E-cadherin; cell adhesion molecule; CAM;  
KW uvomorulin; L-CAM; Cell CAM 120/80.  
OS Homo sapiens.  
PN MO9411401-A.  
PD 26-MAY-1994.  
PF 16-NOV-1993; U11097.  
PR 17-NOV-1992; US-978897.  
PA (UYXA) UNIV YALE.  
PI Morrow JS, Rimm DL;  
DR WPI: 94-183426/22.  
DR N-PSDB: Q65487.  
PT Purified human E-cadherin protein and nucleic acid - used to  
PT develop prods. for diagnosis, prognosis, therapy and prophylaxis  
PT of t-cadherin disorders, e.g. malignancies  
PS Claim 1: Page 59-63; 97pp; English.  
CC E-cadherin is a cell adhesion molecule that is also known as  
CC uvomorulin, L-CAM and Cell CAM 120/80. The DNA encoding hEC was obt.  
CC by screening normal human liver and hepatocellular carcinoma cDNA  
CC libraries and a colonic epithelial cell cDNA library. The following  
CC sequences are specifically claimed: AAs 1-878; 151-878; 30  
CC sequential AAs from AAs 308-878; AAs 1-150; AAs 178-289; AAs 280-  
CC 401; AAs 402-513; AAs 178-513; AAs 151-703; AAs 1-703; AAs 728-878;  
CC AAs 704-878; nucleotide sequences comprising nucleotide numbers 116-  
CC 2749; 566-2749; 1-1053; 510-2696; 1332-3000; 540-1500; 348-906; 890-  
CC 1648; 384-1208; 641-2046; 685-1336; 880-1661; 1199-1742; 1373-1742;  
CC 1705-2204; 2458-2775; DNA encoding at least 30 AAs selected from  
CC AAs 308-878. The prods. can be used in the diagnosis, prognosis,  
CC therapy and prophylaxis of conditions involving improper E-cadherin  
CC expression. Suitable dosages for i.v. admin. of a protein are  
CC 20-500 mcg/kg body wt.  
SQ Sequence 878 AA;

Query Match 54.9%; Score 39; DB 1; Length 878;  
Best Local Similarity 53.8%; Pred. No. 23;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14  
DB 293 NAAIAVITLSQDP 305

RESULT 10  
R85487  
ID R85487 standard; Protein: 878 AA.  
AC R85487.  
DE 18-MAR-1996 (first entry)  
KW Human E-cadherin precursor.  
KW E-cadherin; T-lymphocyte; alpha-E-beta-7 integrin; cell adhesion;  
KW autoimmune disease; Crohn disease; psoriasis.  
OS Homo sapiens.  
PI Key  
DR WPI: 94-150  
DR Location/Qualifiers  
FT peptide  
FT 1, 150  
FT /label- Sig\_peptide

FT domain 151..702  
FT /label- Extracellular\_domain  
FT /note- "the extracellular domain (amino acids 1-552  
FT of the mature protein) is the preferred  
FT region for generation of peptides of the  
FT invention"  
FT 703..726  
FT domain  
FT /label- Transmembrane\_domain  
FT 727..876  
FT /label- Cytoplasmic\_domain  
FT domain  
FT MO9529693-A1.  
FT 09-NOV-1995.  
FT 03-MAY-1995; U05518.  
FT 03-MAY-1994; US-237919.  
FT (BGHM) BRIGHAM & WOMEN'S HOSPITAL.  
FT Brenner MB, Ceppek KL;  
FT WPI: 95-392921/50.  
FT N-PSDB: T05764.  
FT Inhibiting adhesion of T lymphocytes with E-cadherin - useful for  
FT isolating agents to treat auto-immune diseases e.g. Crohn's disease,  
FT psoriasis, etc  
FT Disclosure: Page 70-75; 103pp; English.  
FT The human E-cadherin protein precursor (R85487) is expressed by  
FT an cDNA clone (T05764) derived from human liver. The extracellular  
FT domain of E-cadherin is used to generate peptides that specifically  
FT bind to heterotypic cognates of E-cadherin and which inhibit adhesion  
FT of intra-epithelial T-lymphocytes to E-cadherin-expressing epithelial  
FT or endothelial cells in vitro or in vivo, thereby modulating  
FT mucosal immune responses. Such peptides are also specifically  
FT reactive with a monoclonal antibody (E4.6 or E6.1) that binds to  
FT E-cadherin and that can inhibit T-cell binding.  
SQ Sequence 878 AA;

Query Match 54.9%; Score 39; DB 1; Length 878;  
Best Local Similarity 53.8%; Pred. No. 23;  
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14  
DB 293 NAAIAVITLSQDP 305

RESULT 11  
R97246  
ID R97246 standard; Protein: 4472 AA.  
AC R97246;  
DE 07-JAN-1997 (first entry)  
DE Virulence gene cluster polypeptide product.  
KW Mutant; adaptation; virulence factor; identification; screening;  
KW vaccine; drugs; infection; treatment.  
OS Salmonella typhimurium.  
PI Key  
DR WPI: 96-287194/29.  
DR Identifying virulence genes in microorganisms - by introducing  
PT mutants with insertion inactivated genes into environment and  
PT retrieval and analysis of mutants  
PS Claim 51: Figure-11; 131pp; English.  
CC A method for identifying a microorganism having a reduced adaptation  
CC to a particular environment comprising the steps of: (1) providing a  
CC plurality of microorganisms each of which is independently mutated by

CC the insertional inactivation of a gene with a nucleic acid comprising  
 CC a unique marker sequence so that each mutant contains a different  
 CC marker sequence, or clones of the said microorganism; (2) providing  
 CC individually a stored sample of each mutant produced by step (1) and  
 CC providing individually stored nucleic acid comprising the unique  
 CC marker sequence from each individual mutant; (3) introducing a  
 CC plurality of mutants produced by step (1) into the said particular  
 CC environment and allowing those microorganisms which are able to do so  
 CC to grow in the said environment; (4) retrieving microorganisms from  
 CC the said environment or a selected part thereof and isolating the  
 CC nucleic acid from the retrieved microorganisms; (5) comparing any  
 CC marker sequences in the nucleic acid isolated in step (4) to the  
 CC unique marker sequence of each individual mutant stored as in step  
 CC (2); and (6) selecting an individual mutant which does not contain any  
 CC of the marker sequences as isolated in step (4). The products and  
 CC methods can be used for identifying virulence genes in microorganisms.  
 CC The mutant microorganisms can be used in vaccines or to screen for  
 CC drugs which reduce virulence or compounds useful for preventing,  
 CC ameliorating or treating infections in animals or plants.

CC Sequence 4472 AA;

Query Match 47.9%; Score 34; DB 1; Length 4472;  
 Best Local Similarity 58.3%; Pred. No. 1.3e+03;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 LNSKIAFRIVSOEP 13  
 DB 2873 LNSKIAFRIVSOEP 2884

RESULT 12  
 W48711

ID W48711 standard; Protein; 2233 AA.

AC W48711;

DT 13-OCT-1998 (first entry)

DE HPIV-3 JS isolate wild-type L protein.

KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;

OS single stranded RNA virus; Mononegavirales.

PN WO9813501-A2.

PD 02-APR-1998.

PF 19-SEP-1997; U16718.

PR 27-SEP-1996; US-026823.

PA (AMCY ) AMERICAN CYANAMID CO.

PI (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PT Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;

DR N-PSDB: V18272.

PT Recombinantly-generated, attenuated, non-segmented, negative-sense,

PT single stranded RNA virus of order Mononegavirales - having

PT attenuating mutation in 3' genomic promoter region and RNA

PT polymerase gene, useful as vaccine to immunise against such virus

PS Disclosure: Page 246-254; 426pp; English.

OS This sequence represents the wild-type L protein from Human parainfluenza

CC virus (HPIV-3) type 3 isolate JS. This sequence is used in a method which

CC involves the isolation of recombinantly-generated, attenuated,

CC non-segmented, negative-sense, single stranded RNA virus of the order

CC Mononegavirales which have at least 1 attenuating mutation in the 3'

CC genomic promoter region and at least 1 attenuating mutation in the RNA

CC polymerase gene. This RNA virus can be used as a vaccine to immunise an

CC individual against such a virus.

CC Sequence 2233 AA;

Query Match 47.9%; Score 34; DB 1; Length 2233;

Best Local Similarity 35.7%; Pred. No. 6.1e+02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 LNSKIAFRIVSOEP 14  
 DB 973 LDRSVLYRIMNOEP 986

RESULT 13

W48712

ID W48712 standard; Protein; 2233 AA.

AC W48712;

DT 13-OCT-1998 (first entry)

DE HPIV-3 FRH1 cp45 vaccine L protein.

KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;

OS single stranded RNA virus; Mononegavirales.

PN WO9813501-A2.

PD 02-APR-1998.

PF 19-SEP-1997; U16718.

PR 27-SEP-1996; US-026823.

PA (AMCY ) AMERICAN CYANAMID CO.

PI (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PT Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;

DR N-PSDB: V18273.

PT Recombinantly-generated, attenuated, non-segmented, negative-sense,

PT single stranded RNA virus of order Mononegavirales - having

PT attenuating mutation in 3' genomic promoter region and RNA

PT polymerase gene, useful as vaccine to immunise against such virus

PS Disclosure: Page 265-273; 426pp; English.

OS This sequence represents the human parainfluenza virus (HPIV-3) type 3

CC vaccine FRH1 cp45 L protein. This sequence is used in a method which

CC involves the isolation of recombinantly-generated, attenuated,

CC non-segmented, negative-sense, single stranded RNA virus of the order

CC Mononegavirales which have at least 1 attenuating mutation in the 3'

CC genomic promoter region and at least 1 attenuating mutation in the RNA

CC individual against such a virus.

CC Sequence 2233 AA;

Query Match 47.9%; Score 34; DB 1; Length 2233;

Best Local Similarity 35.7%; Pred. No. 6.1e+02;

Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

OY 1 LNSKIAFRIVSOEP 14  
 DB 973 LDRSVLYRIMNOEP 986

RESULT 14

W48713

ID W48713 standard; Protein; 2233 AA.

AC W48713;

DT 13-OCT-1998 (first entry)

DE HPIV-3 Vero cp45 vaccine L protein.

KW L protein; attenuation; non-segmented; negative sense; vaccine; immunity;

OS single stranded RNA virus; Mononegavirales.

PN WO9813501-A2.

PD 02-APR-1998.

PF 19-SEP-1997; U16718.

PR 27-SEP-1996; US-026823.

PA (AMCY ) AMERICAN CYANAMID CO.

PI (USSH ) US DEPT HEALTH & HUMAN SERVICES.

PT Murphy BR, Randolph VB, Sidhu MS, Tatem JM, Udem SA;

DR N-PSDB: V18274.

PT Recombinantly-generated, attenuated, non-segmented, negative-sense,

PT single stranded RNA virus of order Mononegavirales - having

PT attenuating mutation in 3' genomic promoter region and RNA

PT polymerase gene, useful as vaccine to immunise against such virus

PS Disclosure: Page 283-291; 426pp; English.

OS This sequence represents the Human parainfluenza virus (HPIV-3) type 3

CC vaccine Vero cp45 L protein. This sequence is used in a method which

CC involves the isolation of recombinantly-generated, attenuated,

CC non-segmented, negative-sense, single stranded RNA virus of the order

CC Mononegavirales which have at least 1 attenuating mutation in the 3'

CC genomic promoter region and at least 1 attenuating mutation in the RNA

CC polymerase gene. This RNA virus can be used as a vaccine to immunise an  
 CC individual against such a virus.  
 SO Sequence 2233 AA;

Query Match 47.9%; Score 34; DB 1; Length 2233;  
 Best Local Similarity 35.7%; Pred. No. 6.1e+02;  
 Matches 5; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 LNSKIAFKIVSOEP 14  
 ID 973 LDRSVLYRIMNOEP 986

## RESULT 15

W82841  
 ID W82841 standard; Protein; 456 AA.  
 AC W82841;  
 DT 03-FEB-1999 (first entry)  
 DE Human cerebral protein-1.  
 KW Human; cerebral protein-1; HUCEP-1; Ischaemic brain disease;  
 KW neuron function activating activity; nerve denatured disease;  
 KW Parkinson's disease; Alzheimer's disease.  
 OS Homo sapiens.  
 PN J10257891-A.  
 PD 29-SEP-1998.  
 PR 19-MAR-1997; 065716.  
 PA (TALS) TALSIO PHARM CO LTD.  
 DR WPI; 98-575902/49.  
 DR N-PSDB; V64162.  
 PT HUCEP-1 protein having neuron function activating activity - useful  
 PT for treating ischaemic brain diseases and nerve denaturation  
 PT conditions such as Parkinson's and Alzheimer's diseases  
 PS Claim 1; Fig 4; 1/PP; Japanese.  
 CC The present sequence represents human cerebral protein-1 (HUCEP-1).  
 CC HUCEP-1 has neuron function activating activity. HUCEP-1 is useful  
 CC for treating ischaemic brain diseases and nerve denaturation  
 CC conditions such as Parkinson's and Alzheimer's diseases.  
 SO Sequence 456 AA;

Query Match 47.9%; Score 34; DB 1; Length 456;  
 Best Local Similarity 34.5%; Pred. NO. 98;  
 Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 SKIARIVSOE 13  
 ID 408 TRIGFRIVSKD 418

Search completed: January 6, 2000, 15:40:36  
 Job time: 1683 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 08:52:39 ; Search time 65.16 Seconds  
(without alignments)  
10.858 Million cell updates/sec

Title: US-08-991-628-3  
Perfect score: 71  
Sequence: 1 LNSKIARFIVSQEPA 15

Scoring table: BLOSUM62

Searched: 142080 seqs, 47169319 residues

Database: PIR-62.\*

Word size: 0

Number of hits that pass the threshold: 142080

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	71	100.0	999	1	IJHUG3	desmoglein 3 precu
2	62	87.3	1049	1	IJHUG1	desmoglein 1 precu
3	62	87.3	1043	1	IJBGG1	desmoglein 1 precu
4	55	77.5	1117	2	S38673	desmoglein 2 - hum
5	42	59.2	884	1	IJMSCE	E-cadherin precurs
6	42	59.2	884	2	S34438	uvomorulin mouse
7	41	57.7	732	1	IJCHCB	B-cadherin precurs
8	40	56.3	5147	1	IJFFPM	cadherin-related t
9	40	56.3	1421	2	T02501	hypothetical prote
10	39	54.9	906	1	IJHUCN	cadherin 2 precurs
11	39	54.9	882	1	IJHUCN	cadherin 1 precurs
12	39	54.9	726	1	S73915	virulence-associat
13	39	54.9	829	2	I46356	ksp-cadherin - rab
14	38	53.5	906	1	IJXLC2	N-cadherin 2 precu
15	38	53.5	905	1	IJXLC1	N-cadherin 1 precu
16	38	53.5	905	2	S43064	cadherin - African
17	38	53.5	385	2	S49752	hypothetical prote
18	38	53.5	138	2	S74520	hypothetical prote
19	38	53.5	333	2	JC5040	positive regulator
20	38	53.5	533	2	S57904	virA9 protein - S
21	37	52.1	493	1	S39532	aldehyde dehydroge
22	37	52.1	236	2	D71733	phosphoribosylamin
23	37	52.1	461	2	A71662	sodium/pantothenat
24	37	52.1	135	2	H72471	hypothetical prote
25	36	50.7	877	1	IJBGCN	N-cadherin precurs
26	36	50.7	906	1	IJMSCN	N-cadherin precurs
27	36	50.7	887	1	IJHUC1	E-cadherin precurs
28	36	50.7	829	1	IJHUCP	cadherin 3 precurs
29	36	50.7	822	1	IJMSCP	P-cadherin precurs
30	36	50.7	725	1	E64211	virulence-associat
31	36	50.7	276	2	JC5285	carboxyl reductase
32	36	50.7	277	2	JC5284	carboxyl reductase
33	36	50.7	475	2	H71660	nitrogen assimilati
34	36	50.7	742	2	A49341	isocitrate dehydro
35	36	50.7	103	2	D70897	hypothetical prote

36 36 50.7 182 2 H64386 hypothetical prote  
37 36 50.7 1264 2 S64146 probable membrane  
38 36 50.7 2122 2 B75009 ribonucleotide red  
39 35 49.3 342 1 CFPM plastoquinol--plas  
40 35 49.3 320 1 CFPN plastoquinol--plas  
41 35 49.3 320 1 CFPN plastoquinol--plas  
42 35 49.3 643 1 S24382 nitrous-oxide redu  
43 35 49.3 2228 1 ZLN2SV genome polypeptid  
44 35 49.3 2048 1 ZLN2SE genome polypeptid  
45 35 49.3 241 2 S63634 ribosomal protein

## ALIGNMENTS

RESULT 1  
IJHUG3  
desmoglein 3 precursor - human  
N:Alternate names: pemphigus vulgaris antigen  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence: revision 30-Jun-1993 #text\_change 22-Jun-1999  
R:Amagai, M.; Klaus-Kovtun, V.; Stanley, J.R.  
Cell 67, 869-877, 1991  
A:Title: Autoantibodies against a novel epithelial cadherin in pemphigus vulgaris, a  
A:Reference number: A41088, PMID:92069753  
A:Accession: A41088  
A:Molecule type: mRNA  
A:Residues: 1-999 <AMA>  
A:Cross-references: GB:W76482; NID:g190751; PIDN:AAA60230.1; PID:g190752  
C:Genetics:  
A:Gene: GDB:DSG3  
A:Cross-references: GDB:134030; OMIM:169615  
A:Map position: 18q12.1-18q12.2  
C:Superfamily: cadherin; cadherin repeat homology  
C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-49/Domain: propeptide #status predicted <PRO>  
F:50-999/Product: desmoglein homolog #status predicted <MAT>  
F:50-999/Domain: extracellular #status predicted <EXT>  
F:52-157/Domain: cadherin repeat homology <CR1>  
F:160-267/Domain: cadherin repeat homology <CR2>  
F:270-383/Domain: cadherin repeat homology <CR3>  
F:390-495/Domain: cadherin repeat homology <CR4>  
F:496-598/Domain: cadherin repeat homology <CR5>  
F:616-939/Domain: transmembrane #status predicted <TM>  
F:640-999/Domain: intracellular #status predicted <INT>  
F:910-938/Domain: desmoglein repeat <DG1>  
F:937-966/Domain: desmoglein repeat <DG2>  
F:110,180,545/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 100.0%; Score 71; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 2,6e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LNSKIARFIVSQEPA 15  
DB 190 LNSKIARFIVSQEPA 204

RESULT 2  
IJHUG1  
desmoglein 1 precursor - human  
N:Alternate names: desmosomal glycoprotein I  
C:Species: Homo sapiens (man)  
C>Date: 30-Jun-1993 #sequence: revision 30-Jun-1993 #text\_change 22-Jun-1999  
C:Accession: S16906; A37066; A61254; A61279; S16158  
R:Buxton, R.S.  
Submitted to the EMBL Data Library, November 1990  
A:Reference number: S16906  
A:Accession: S16906  
A:Molecule type: mRNA

A:Residues: 1-1049 <BUX>  
 A:Cross-references: EMBL:X56654; NID:930505; PIDN:CAA39976.1; PID:930506  
 R:Wheeler, G.N.; Parker, A.E.; Thomas, C.L.; Atalio, P.; Poynter, D.; Arnemann, J.; R  
 Proc. Natl. Acad. Sci. U.S.A. 88, 4796-4800, 1991  
 A:Title: Desmosomal glycoprotein DGI, a component of intercellular desmosome junctions,  
 A:Reference number: A39706; MUID:91271279  
 A:Accession: A39706  
 A:Molecule type: mRNA  
 A:Residues: 24-1049 <WHE>  
 A:Cross-references: GB:X56654  
 R:Nilles, L.A.; Parry, D.A.D.; Powers, E.E.; Angst, B.D.; Wagner, R.M.; Green, K.J.  
 J. Cell Sci. 99, 809-821, 1991  
 A:Title: Structural analysis and expression of human desmoglein: a cadherin-like compone  
 A:Reference number: A61254; MUID:92121251  
 A:Accession: A61254  
 A:Molecule type: mRNA  
 A:Residues: 26-1049 <NLL>  
 R:Wheeler, G.N.; Buxton, R.S.; Parker, A.E.; Arnemann, J.; Rees, D.A.; King, I.A.; Magee  
 Biochem. Soc. Trans. 19, 1060-1064, 1991  
 A:Title: Desmosomal glycoproteins I, II and III: novel members of the cadherin superfamily  
 A:Reference number: A61279; MUID:921515187  
 A:Accession: A61279  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-55 <WR3>  
 C:Genetics:  
 A:Gene: GDB:DSG1  
 A:Cross-references: GDB:126563; OMIM:125670  
 A:Map position: 18q12.1-18q12.2  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-49/Domain: propeptide #status predicted <PRO>  
 F:50-1049/Domain: extracellular #status predicted <EXT>  
 F:50-548/Domain: extracellular #status predicted <EXT>  
 F:52-157/Domain: cadherin repeat homology <CR1>  
 F:160-269/Domain: cadherin repeat homology <CR2>  
 F:272-385/Domain: cadherin repeat homology <CR3>  
 F:392-493/Domain: cadherin repeat homology <CR4>  
 F:509-569/Domain: transmembrane #status predicted <TM>  
 F:549-569/Domain: transmembrane #status predicted <TM>  
 F:572-1049/Domain: intracellular #status predicted <INT>  
 F:840-869/Domain: desmoglein repeat <DG1>  
 F:870-889/Domain: desmoglein repeat <DG2>  
 F:900-927/Domain: desmoglein repeat <DG3>  
 F:928-956/Domain: desmoglein repeat <DG4>  
 F:969-1019/Region: glycine/serine-rich  
 F:110,180/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.3%; Score 62; DB 1; Length 1049;  
 Best Local Similarity 80.0%; Pred. No. 0.0014;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
 Db 190 LNSKIAFKITROEPS 204

RESULT 3  
 IJBOG1  
 desmoglein 1 precursor - bovine  
 N:Alternate names: desmoglein BDGM  
 C:Species: Bos primigenius taurus (cattle)  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999  
 C:Accession: S14603; A38872; A48173; S24412  
 R:Koch, P.J.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.W.  
 submitted to the EMBL Data Library, March 1991  
 A:Description: Complete sequence of the desmoglein precursor and evidence for the existe  
 A:Reference number: S14603  
 A:Accession: S14603  
 A:Molecule type: mRNA  
 A:Residues: 1-1043 <KOC>

A:Cross-references: EMBL:X58466; NID:9306; PIDN:CAA41380.1; PID:9307  
 R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimbelmann, R.; Franke, W.W.  
 Eur. J. Cell Biol. 55, 200-208, 1991  
 A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypept  
 A:Reference number: A38872; MUID:92037656  
 A:Accession: A38872  
 A:Molecule type: mRNA  
 A:Residues: 1-87,968-1043 <KO2>  
 A:Cross-references: GB:S64268; GB:S64270  
 R:Goodwin, L.; Hill, J.E.; Raynor, K.; Raszi, L.; Manabe, M.; Cowin, P.  
 Biochem. Biophys. Res. Commun. 173, 1224-1230, 1990  
 A:Title: Desmoglein shows extensive homology to the cadherin family of cell adhesion  
 A:Reference number: A37785; MUID:91097553  
 A:Accession: A37785  
 A:Molecule type: mRNA  
 A:Residues: 44-123, 'V', 125-493 <GOO>  
 A:Cross-references: GB:M58165; NID:9162966; PIDN:AAA62709.1; PID:9552318  
 R:Zimbelmann, R.  
 submitted to the EMBL Data Library, February 1991  
 A:Reference number: S38721  
 A:Accession: S38721  
 A:Molecule type: mRNA  
 A:Residues: 44-1043 <ZIX>  
 A:Cross-references: EMBL:X57784; NID:9436061; PIDN:CAA40930.1; PID:9436062  
 R:Koch, P.J.; Walsh, M.J.; Schmelz, M.; Goldschmidt, M.D.; Zimbelmann, R.; Franke, W.  
 Eur. J. Cell Biol. 53, 1-12, 1990  
 A:Title: Identification of desmoglein, a constitutive desmosomal glycoprotein, as a m  
 A:Reference number: A48173; MUID:91168965  
 A:Accession: A48173  
 A:Molecule type: mRNA  
 A:Residues: 44-1001, 'AOPPSAT' <KO3>  
 A:Cross-references: GB:X57784  
 A:Note: this sequence has been revised in references A38872 and S38721  
 C:Genetics:  
 A:Gene: DSG1  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane  
 F:1-23/Domain: signal sequence #status predicted <SIG>  
 F:24-49/Domain: propeptide #status predicted <PRO>  
 F:50-1043/Domain: extracellular #status predicted <EXT>  
 F:50-548/Domain: extracellular #status predicted <EXT>  
 F:52-157/Domain: cadherin repeat homology <CR1>  
 F:160-269/Domain: cadherin repeat homology <CR2>  
 F:272-385/Domain: cadherin repeat homology <CR3>  
 F:392-491/Domain: cadherin repeat homology <CR4>  
 F:549-574/Domain: transmembrane #status predicted <TM>  
 F:575-1043/Domain: intracellular #status predicted <INT>  
 F:846-875/Domain: desmoglein repeat <DG1>  
 F:876-905/Domain: desmoglein repeat <DG2>  
 F:906-933/Domain: desmoglein repeat <DG3>  
 F:934-962/Domain: desmoglein repeat <DG4>  
 F:963-1012/Region: glycine/serine-rich  
 F:110/Binding site: carbohydrate (Asn) (covalent) #status experimental  
 F:180,496/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 87.3%; Score 62; DB 1; Length 1043;  
 Best Local Similarity 80.0%; Pred. No. 0.0014;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
 Db 190 LNSKIAFKITROEPS 204

RESULT 4  
 S38673  
 desmoglein 2 - human  
 N:Alternate names: desmoglein HDGC  
 C:Species: Homo sapiens (man)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 20-Aug-1999  
 C:Accession: S38673; B38872  
 R:Zimbelmann, R.

submitted to the EMBL Data Library, September 1993

A:Reference number: S38673

A:Accession: S38673

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1117 <21M>

A:Cross-references: EMBL:Z26317; NID:9416177; PIDN:CA81226.1; PID:9416178

R:Koch, P.J.; Goldschmidt, M.D.; Walsh, M.J.; Zimblemann, R.; Franke, W.W.

Eur. J. Cell Biol. 55, 200-208, 1991

A:Title: Complete amino acid sequence of the epidermal desmoglein precursor polypeptide

A:Reference number: A38672; MUID:92037656

A:Accession: B38672

A:Molecule type: mRNA

A:Residues: 777-1117 <KOC>

A:Cross-references: GB:S64273

C:Genetics:

A:Gene: GDB:DSG2

A:Cross-references: GDB:128808; OMIM:125671

A:Map position: 18q12.1-18q12.2

C:Superfamily: cadherin; cadherin repeat homology

C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; membrane protein

F:51-158/Domain: cadherin repeat homology <CR1>

F:161-271/Domain: cadherin repeat homology <CR2>

Query Match 77.5%; Score 55; DB 2; Length 1117;

Best Local Similarity 73.3%; Pred. No. 0.032;

Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15

DB 191 LNSKISRYVSLPEA 205

RESULT 5

1JMSCE

E-cadherin precursor, epithelial - mouse

N:Alternate names: uvomorulin

C:Species: Mus musculus (house mouse)

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999

C:Accession: S04528; S03160; I49565; S48735

R:Nagafochi, A.; Shirayoshi, Y.; Okazaki, K.; Yasuda, K.; Takeichi, M.

Nature 339, 341-343, 1987

A:Title: Transformation of cell adhesion properties by exogenously introduced E-cadherin

A:Reference number: S04528; MUID:87315445

A:Accession: S04528

A:Molecule type: mRNA

A:Residues: 1-412, 'V', 414-884 <NAG>

A:Cross-references: EMBL:X06115

R:Ringwald, M.; Schuh, R.; Vestweber, D.; Eistetter, H.; Lottspeich, F.; Engel, J.; Doel

EMBO J. 6, 3647-3653, 1987

A:Title: The structure of cell adhesion molecule uvomorulin. Insights into the molecular

A:Reference number: S03160; MUID:88111553

A:Accession: S03160

A:Molecule type: mRNA

A:Residues: 157-884 <RIN>

A:Cross-references: EMBL:X06339

A:Note: part of this sequence, including the amino end of the mature protein, was confir

R:Behrens, J.; Loewerick, O.; Klein-Hitpass, L.; Birchmeier, W.

Proc. Natl. Acad. Sci. U.S.A. 88, 11495-11499, 1991

A:Title: The E-cadherin promoter: Functional analysis of a G-C-rich region and an epithe

A:Reference number: I49565; MUID:92107977

A:Accession: I49565

A:Status: preliminary; translated from GB/EMBL/DBD1

A:Molecule type: DNA

A:Residues: 1-15 <RES>

A:Cross-references: GB:M81449; NID:9192325; PIDN:AAA37352.1; PID:9192326

R:Hong, K.I.; Yau, P.; Overduin, M.; Bagby, S.; Portumb, T.; Takeichi, M.; Ikura, M.

FEBS Lett. 359, 318-322, 1994

A:Title: Purification and spectroscopic characterization of a recombinant amino-terminal

A:Reference number: S48735; MUID:95010732

A:Accession: S48735

A:Status: preliminary

A:Molecule type: protein

A:Residues: 156-300 <TON>

C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are though

C:Genetics:

A:Gene: E-cadherin

C:Superfamily: cadherin; cadherin repeat homology

C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane

F:1-27/Domain: signal sequence #status predicted <PRO>

F:127/Domain: signal sequence #status predicted <SIG>

F:157-884/Product: E-cadherin, epithelial #status experimental <MAT>

F:157-884/Domain: extracellular #status predicted <EXT>

F:159-264/Domain: cadherin repeat homology <CR1>

F:234-239/Region: cadherin binding #status predicted

F:267-377/Domain: cadherin repeat homology <CR2>

F:380-488/Domain: cadherin repeat homology <CR3>

F:489-597/Domain: cadherin repeat homology <CR4>

F:598-702/Domain: cadherin repeat homology <CR5>

F:707-733/Domain: transmembrane #status predicted <TM>

F:734-884/Domain: intracellular #status predicted <INT>

F:842-855/Region: serine-rich

F:560,639/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 59.2%; Score 42; DB 1; Length 884;

Best Local Similarity 61.5%; Pred. No. 7.1;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14

DB 299 NNAIAIVTVSQDP 311

RESULT 6

S34438

uvomorulin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 22-Nov-1995 #sequence\_revision 03-Nov-1995 #text\_change 23-May-1997

C:Accession: S34438

R:Ringwald, M.; Baribault, H.; Schmidt, C.; Kemler, R.

Nucleic Acids Res. 19, 6533-6539, 1991

A:Title: The structure of the gene coding for the mouse cell adhesion molecule uvomor

A:Reference number: S34438; MUID:92093614

A:Accession: S34438

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Cross-references: EMBL:X06975

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992

C:Genetics:

A:Insertions: 16/3; 57/1; 131/3; 179/3; 231/3; 280/1; 338/3; 381/3; 442/3; 524/2; 573/1

C:Superfamily: cadherin; cadherin repeat homology

F:380-488/Domain: cadherin repeat homology <CR3>

Query Match 59.2%; Score 42; DB 2; Length 884;

Best Local Similarity 61.5%; Pred. No. 7.1;

Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14

DB 299 NNAIAIVTVSQDP 311

RESULT 7

IICHCB

B-cadherin precursor - chicken (fragment)

N:Alternate names: K-CAM protein

C:Species: Gallus gallus (chicken)

C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 22-Jun-1999

C:Accession: A41634; A48715; S16160

R:Stork, B.C.; Gallin, W.J.; Edelman, G.M.; Cunningham, B.A.

Proc. Natl. Acad. Sci. U.S.A. 88, 11545-11549, 1991

A:Title: Genes for two calcium-dependent cell adhesion molecules have similar structu

A:Reference number: A41634; MUID:92107987  
 A:Accession: A41634  
 A:Molecule type: DNA  
 A:Residues: 1-732 <SOR>  
 A:Cross-references: GB:841894; NID:9212226; PIDN:AAA48929.1; PID:9212227  
 J: Napolitano, E.W.; Venstrom, K.; Wheeler, E.F.; Reichardt, L.F.  
 R: Cell Biol. 113: 893-905, 1991  
 A:Title: Molecular cloning and characterization of B-cadherin, a novel chick cadherin.  
 A:Reference number: A38715; MUID:91225083  
 A:Accession: A38715  
 A:Molecule type: mRNA  
 A:Residues: 7-413; 'V', 445-732 <NAP>  
 A:Cross-references: GB:X58518; NID:963113; PIDN:CAA41408.1; PID:963114  
 C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to  
 C:Genetics:  
 A:Gene: K-CAM  
 A:Introns: 29/3; 81/3; 130/1; 188/3; 231/3; 293/3; 375/2; 423/1; 498/1; 571/1; 614/3; 66  
 C:Superfamily: cadherin; cadherin repeat homology  
 C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pr  
 F:1-6/Domain: propeptide (fragment) #status predicted <EXT>  
 F:6-554/Domain: extracellular #status predicted <MAT>  
 F:7-732/Product: B-cadherin #status predicted <MAT>  
 F:9-114/Domain: cadherin repeat homology <CR1>  
 F:84-89/Region: cadherin binding #status predicted  
 F:117-227/Domain: cadherin repeat homology <CR2>  
 F:230-339/Domain: cadherin repeat homology <CR3>  
 F:340-447/Domain: cadherin repeat homology <CR4>  
 F:448-550/Domain: cadherin repeat homology <CR5>  
 F:555-580/Domain: transmembrane #status predicted <TM>  
 F:581-732/Domain: intracellular #status predicted <INT>  
 F:689-702/Region: serine-rich  
 F:137,410/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 57.7%; Score 41; DB 1; Length 732;  
 Best Local Similarity 61.5%; Pred. No. 8.9;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 NSKIAFKVISOEP 14  
 1 11: 11111  
 Db 149 NGVAVYSILSOEP 161

RESULT 8  
 IUFFM  
 cadherin-related tumor suppressor precursor - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Feb-1997  
 C:Accession: A41087; B41087  
 R:Mahoney, P.A.; Weber, U.; Onofrechuk, P.; Biesmann, H.; Bryant, P.J.; Goodman, C.S.  
 Cell 67: 853-868, 1991  
 A:Title: The fat tumor suppressor gene in Drosophila encodes a novel member of the cadhe  
 A:Reference number: A41087; MUID:92069752  
 A:Accession: A41087  
 A:Molecule type: mRNA  
 A:Residues: 143-485; 1279-5147 <MAH>  
 A:Cross-references: GB:M60537  
 A:Accession: B41087  
 A:Molecule type: DNA  
 A:Residues: 1-142; 487-1278 <MA2>  
 A:Cross-references: GB:M60537  
 A:Note: 1229-Gly and 1233-Set were also found  
 C:Genetics:  
 A:Gene: fat  
 A:Cross-references: FlyBase:FBgn0001075  
 C:Superfamily: cadherin-related tumor suppressor; cadherin repeat homology; EGF homology  
 C:Keywords: calcium binding; cell adhesion; duplication; transmembrane protein  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:36-5147/Product: cadherin-related tumor suppressor #status predicted <MAT>  
 F:36-4383/Domain: extracellular #status predicted <EXT>  
 F:51-156/Domain: cadherin repeat homology <CR1>  
 F:159-270/Domain: cadherin repeat homology <CR2>  
 F:271-383/Domain: cadherin repeat homology <CR3>

F:390-494/Domain: cadherin repeat homology <CR4>  
 F:497-599/Domain: cadherin repeat homology <CR5>  
 F:602-708/Domain: cadherin repeat homology <CR6>  
 F:718-822/Domain: cadherin repeat homology <CR7>  
 F:831-942/Domain: cadherin repeat homology <CR8>  
 F:948-1049/Domain: cadherin repeat homology <CR9>  
 F:1052-1153/Domain: cadherin repeat homology <CR10>  
 F:1156-1278/Domain: cadherin repeat homology <CR11>  
 F:1281-1384/Domain: cadherin repeat homology <CR12>  
 F:1387-1489/Domain: cadherin repeat homology <CR13>  
 F:1492-1601/Domain: cadherin repeat homology <CR14>  
 F:1607-1713/Domain: cadherin repeat homology <CR15>  
 F:1717-1823/Domain: cadherin repeat homology <CR16>  
 F:1826-1922/Domain: cadherin repeat homology <CR17>  
 F:1925-2027/Domain: cadherin repeat homology <CR18>  
 F:2028-2167/Domain: cadherin repeat homology <CR19>  
 F:2169-2278/Domain: cadherin repeat homology <CR20>  
 F:2281-2384/Domain: cadherin repeat homology <CR21>  
 F:2387-2491/Domain: cadherin repeat homology <CR22>  
 F:2494-2596/Domain: cadherin repeat homology <CR23>  
 F:2599-2703/Domain: cadherin repeat homology <CR24>  
 F:2707-2810/Domain: cadherin repeat homology <CR25>  
 F:2813-2913/Domain: cadherin repeat homology <CR26>  
 F:2915-3013/Domain: cadherin repeat homology <CR27>  
 F:3014-3124/Domain: cadherin repeat homology <CR28>  
 F:3127-3229/Domain: cadherin repeat homology <CR29>  
 F:3232-3334/Domain: cadherin repeat homology <CR30>  
 F:3337-3439/Domain: cadherin repeat homology <CR31>  
 F:3442-3545/Domain: cadherin repeat homology <CR32>  
 F:3548-3651/Domain: cadherin repeat homology <CR33>  
 F:3654-3756/Domain: cadherin repeat homology <CR34>  
 F:3759-3861/Domain: EGF homology <EG1>  
 F:4017-4048/Domain: EGF homology <EG2>  
 F:4056-4089/Domain: EGF homology <EG3>  
 F:4096-4127/Domain: EGF homology <EG4>  
 F:484-4609/Domain: transmembrane #status predicted <TM>  
 F:4610-5147/Domain: intracellular #status predicted <INT>

Query Match 56.3%; Score 40; DB 1; Length 5147;  
 Best Local Similarity 50.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LNSKIAFKVISOEP 14  
 1 11: 11111  
 Db 1520 LNSKIAFKVISOEP 1533

RESULT 9  
 T02501  
 hypothetical protein T19C21.7 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 14-May-1999  
 C:Accession: T02501  
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
 submitted to the EMBL data library/ August 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.  
 A:Reference number: Z14676  
 A:Accession: T02501  
 A:Status: translated from GB/EMBL/DDBAT  
 A:Molecule type: DNA  
 A:Residues: 1-1421 <ROU>  
 A:Cross-references: EMBL:AC004683; NID:93395421; PID:93395428  
 A:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 2  
 A:Introns: 52/2; 107/1; 148/2; 191/3; 212/2; 1317/3; 1346/3  
 A:Note: T19C21.7

Query Match 56.3%; Score 40; DB 2; Length 1421;  
 Best Local Similarity 42.9%; Pred. No. 29;  
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEP 14  
 Db 262 VDSKISYEITITOMP 275

## RESULT 10

IUHUCN

cadherin 2 precursor - human  
 N.Alternate names: N-cadherin; neuronal cadherin

C.Species: Homo sapiens (man)  
 C.Date: 30-Jun-1993 #sequence, revision 30-Jun-1993 #text, change 22-Jun-1999

C.Accession: A38870; S11487; J00751; S13799

R.Reid, R.A.

submitted to the EMBL Data Library, November 1990

A.Reference number: A38870

A.Accession: A38870

A.Molecule type: mRNA

A.Residues: 1-906 <RE1>  
 A.Cross-references: EMBL:X54315; NID:934998; PIDN:CAA38213.1; PID:934999

R.Reid, R.A.; Hemperly, J.J.

Nucleic Acids Res. 18, 5896, 1990

A>Title: Human N-cadherin: nucleotide and deduced amino acid sequence.

A.Reference number: S11487; MUID:91016946

A.Accession: S11487

A.Molecule type: mRNA

A.Residues: 1-340, 'N', '342-698', 'R', '700-704', 'F', '706-906 <RE2>

A.Cross-references: EMBL:X54315

A>Note: this sequence has been revised in reference A38870

R.Walsh, F.S.; Barton, C.H.; Putt, W.; Moore, S.E.; Kelsell, D.; Spurr, N.; Goodfellow, J. Neurochem. 55, 805-812, 1990

A>Title: N-cadherin gene maps to human chromosome 18 and is not linked to the E-cadherin

A.Reference number: J00751; MUID:90347462

A.Accession: J00751

A.Molecule type: mRNA

A.Residues: 160-194, 'IR', '197-211', 'L', '213-227', 'Q', '229', 'N', '231-235', 'G', '237-248', 'T', '250-356

A.Cross-references: GB:W344064

A.Comment: Cadherins mediate calcium-dependent intercellular adhesion and are thought to

C.Genetics:

A.Gene: GDB:CDH2; NCAD

A.Cross-references: GDB:128185; OMIM:114020

A.Map position: 18q12.1

C.Superfamily: cadherin; cadherin repeat homology

C.Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro

F:1-27/Domain: signal sequence #status predicted <SIG>

F:28-159/Domain: propeptide #status predicted <PRO>

F:160-906/Product: N-cadherin #status predicted <EXT>

F:160-714/Domain: extracellular #status predicted <EXT>

F:162-267/Domain: cadherin binding #status predicted

F:237-242/Region: cadherin binding #status predicted

F:270-382/Domain: cadherin repeat homology <CR2>

F:385-487/Domain: cadherin repeat homology <CR3>

F:500-605/Domain: cadherin repeat homology <CR4>

F:606-716/Domain: cadherin repeat homology <CR5>

F:715-746/Domain: transmembrane #status predicted <TM>

F:747-906/Domain: intracellular #status predicted <INT>

F:865-878/Region: serine-rich

F:190,273,325,402,572,622,651,692/Binding site: carbohydrate (Asn) (covalent) #status pr

Query Match 54.9%; Score 39; DB 1; Length 906;

Best Local Similarity 46.7%; Pred. No. 27;

Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEP 15

Db 300 LNSKIAFKIVSOAPS 314

## RESULT 11

IUHUCN

cadherin 1 precursor - human  
 N.Alternate names: ARC-1; cell CAM 120/80; E-cadherin; epithelial cadherin; L-CAM; uvomor

C.Species: Homo sapiens (man)

C.Date: 30-Jun-1993 #sequence, revision 30-Jun-1993 #text, change 22-Jun-1999

C.Accession: S37654; S31430; S05475; S31460; S06716; A57171; J02230; I52294; I52704;

R.Bussemaekers, M.J.G.; van Bokhoven, A.; Mees, S.G.M.; Kemler, R.; Schalken, J.A.

Mol. Biol. Rep. 17, 123-128, 1993

A>Title: Molecular cloning and characterization of the human E-cadherin cDNA.

A.Reference number: S37654; MUID:93211394

A.Accession: S37654

A.Molecule type: mRNA

A.Residues: 1-882 <BUS>

A.Cross-references: EMBL:Z13009; NID:931072; PIDN:CAA78353.1; PID:931073

R.Keller, W.; Warda, A.; Oda, T.; Hirohashi, S.; Kemler, R.; Birchmeier, W.

submitted to the EMBL Data Library, December 1992

A.Description: Sequence of human E-cadherin cDNA.

A.Reference number: S31430

A.Accession: S31430

A.Molecule type: mRNA

A.Residues: 1-542, 'F', '544-882 <REL>

A.Cross-references: EMBL:Z18923; NID:931074; PIDN:CAA79356.1; PID:931075

R.Hanscourt, A.; Spurr, N.; Goodfellow, P.N.; Kemler, R.

Differentiation 38, 67-71, 1998

A>Title: Characterization and chromosomal localization of the gene encoding the human

A.Reference number: S05475; MUID:89031725

A.Accession: S05475

A.Molecule type: mRNA

A.Residues: 157-311 <MAN>

A.Cross-references: EMBL:X12790

A>Note: nucleotide sequence is not complete

R.Frixen, U.H.

submitted to the EMBL Data Library, March 1990

A.Reference number: S31460

A.Accession: S31460

A.Molecule type: mRNA

A.Residues: 265-392 <FR1>

A.Cross-references: EMBL:X52279; NID:928821; PIDN:CAA36522.1; PID:928822

J.Wheeler, M.J.; Buck, C.A.; Bechtol, K.B.; Damsky, C.H.

J. Cell. Biochem. 34, 187-202, 1987

A>Title: Soluble 80-kd fragment of cell-CAM 120/80 disrupts cell-cell adhesion.

A.Reference number: S06716; MUID:87280410

A.Accession: S06716

A.Molecule type: protein

A.Residues: 'XQ', '157-162', 'V', '164-179 <WB>

R.Berx, G.; Staes, K.; Van Hengel, J.; Molemans, F.; Bussemakers, M.J.G.; van Bokhoven

Genomics 25, 281-289, 1995

A>Title: Cloning and characterization of the human invasion suppressor gene E-cadherin

A.Reference number: A57171; MUID:95324920

A.Accession: A57171

A.Molecule type: DNA

A.Residues: 1-30,32-882 <BR>

A.Cross-references: GB:U34784

R.Rimm, D.L.; Morrow, J.S.

Biochem. Biophys. Res. Commun. 200, 1754-1761, 1994

A>Title: Molecular cloning of human E-cadherin suggests a novel subdivision of the c

A.Reference number: J02230; MUID:94242050

A.Accession: J02230

A.Molecule type: mRNA

A.Residues: 1-9, 'G', '11-15', 'RSPGSGERSPPCLTRELHVGAPAPPEKPR', '52-67', 'I', '69', 'LTPP', '71

A.Cross-references: GB:U08599; NID:9340184; PIDN:AAA61259.1; PID:9340185

A>Note: the majority of differences between this and other reports represent apparan

R.Bussemaekers, M.J.G.; Giaroli, L.A.; van Bokhoven, A.; Schalken, J.A.

Biochem. Biophys. Res. Commun. 203, 1284-1290, 1994

A>Title: Transcriptional regulation of the human E-cadherin gene in human prostate c

A.Reference number: I52294; MUID:94380041

A.Accession: I52294

A>Status: translation not shown; translated from GB/EMBL/DBJ

A.Molecule type: DNA

A.Residues: 1-16 <RES>

A.Cross-references: GB:U34545; NID:9509604; PIDN:AAA21764.1; PID:9509605

R.Becker, K.F.; Ackinson, M.J.; Reich, U.; Becker, I.; Nekarda, H.; Stewert, J.R.; H

Cancer Res. 54, 3845-3852, 1994

A>Title: E-cadherin gene mutations provide clues to diffuse type gastric carcinomas.





R:Detrick, R.J.; Dickey, D.; Kintner, C.R.  
Neuron 4, 493-506, 1990

A:Title: The effects of N-cadherin misexpression on morphogenesis in Xenopus embryos.  
A:Reference number: J00442; MUID:90211966

A:Accession: J00442  
A:Molecule type: mRNA

A:Residues: 1-905 <DET>

C:Comment: Cadherins mediate calcium-dependent intercellular adhesion, and are thought to

C:Superfamily: cadherin; cadherin repeat homology

C:Keywords: calcium binding; cell adhesion; duplication; glycoprotein; transmembrane pro

F:1-28/Domain: signal sequence #status predicted <SIG>

F:29-160/Domain: propeptide #status predicted <PRO>

F:161-905/Product: N-cadherin 1 #status predicted <MAT>

F:161-713/Domain: extracellular #status predicted <EXT>

F:163-268/Domain: cadherin repeat homology <CR1>

F:238-243/Region: cadherin binding #status predicted

F:271-383/Domain: cadherin repeat homology <CR2>

F:386-498/Domain: cadherin repeat homology <CR3>

F:501-606/Domain: cadherin repeat homology <CR4>

F:607-713/Domain: cadherin repeat homology <CR5>

F:714-745/Domain: transmembrane #status predicted <TM>

F:746-905/Domain: intracellular #status predicted <INT>

F:862-877/Region: serine-rich

F:191,274,326,403,573,623,651,692/Binding site: carbohydrate (asn) (covalent) #status pr

#### Query Match

Best Local Similarity 53.5%; Score 38; DB 1; Length 905;  
Pred. No. 42;

Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 NSKIARIVSQEPA 15  
DB 302 NGMLRYKILSQTPA 315

Search completed: January 7, 2000, 08:52:41  
Job time: 2271 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2000, 13:25:37; Search time 28.55 Seconds  
(without alignments)  
15,282 Million cell updates/sec

Title: US-08-991-628-3  
Perfect score: 71  
Sequence: 1 LNSKIAKRVSOEPA 15

Scoring table: BLOSUM62

Searched: 80000 seqs, 29085965 residues

Database: SwissProt\_38.\*

Word size: 0

Number of hits that pass the threshold: 80000

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	71	100.0	999	1	DSG3_HUMAN
2	62	87.3	1043	1	DSG1_BOVIN
3	62	87.3	1049	1	DSG1_HUMAN
4	55	77.5	1117	1	DSG2_HUMAN
5	42	59.2	884	1	CAD1_MOUSE
6	41	57.7	732	1	CADB_CHICK
7	40	56.3	5147	1	FAT_DROME
8	39	54.9	882	1	CAD1_HUMAN
9	39	54.9	906	1	CAD2_HUMAN
10	39	54.9	906	1	VACB_MYCPN
11	38	53.5	925	1	CADB_XENLA
12	38	53.5	905	1	CADN_XENLA
13	38	53.5	906	1	CADO_XENLA
14	38	53.5	385	1	YOK1_YEAST
15	37	52.1	493	1	DHAG_HUMAN
16	36	50.7	887	1	CAD1_CHICK
17	36	50.7	877	1	CAD2_BOVIN
18	36	50.7	906	1	CAD2_MOUSE
19	36	50.7	822	1	CAD3_HUMAN
20	36	50.7	822	1	CAD3_MOUSE
21	36	50.7	276	1	DHCA_RAT
22	36	50.7	741	1	IDR2_VITB1
23	36	50.7	725	1	VACB_MYCGE
24	36	50.7	182	1	Y696_METYA
25	36	50.7	1264	1	YGN3_YEAST
26	35	49.3	1146	1	AS10_YEAST
27	35	49.3	320	1	CYF_MAIZE
28	35	49.3	318	1	CYF_OENHO
29	35	49.3	320	1	CYF_ORISA
30	35	49.3	320	1	CYF_PEA
31	35	49.3	320	1	CYF_PORPU
32	35	49.3	320	1	CYF_POBAC
33	35	49.3	320	1	CYF_VICFA
34	35	49.3	320	1	CYF_WHEAT
35	35	49.3	1419	1	MDR_PLAUF
36	35	49.3	643	1	NOS2_ALCEU
37	35	49.3	2228	1	RRPL_SENDS
38	35	49.3	2048	1	RRPL_SENDE
39	35	49.3	2228	1	RRPL_SENDF

## ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	999 AA
1	DSG3_HUMAN	DSG3_HUMAN	STANDARD	PRT	999 AA
40	35	49.3	2228	1	RRPL_SENDF
41	35	49.3	809	1	TER1_CAEEL
42	35	49.3	760	1	VACB_SHITL
43	34	47.9	541	1	ASNH_METYA
44	34	47.9	633	1	BEM4_YEAST
45	34	47.9	367	1	YENK_CAEEL
46	34	47.9	367	1	YENK_CAEEL
47	34	47.9	367	1	YENK_CAEEL
48	34	47.9	367	1	YENK_CAEEL
49	34	47.9	367	1	YENK_CAEEL
50	34	47.9	367	1	YENK_CAEEL
51	34	47.9	367	1	YENK_CAEEL
52	34	47.9	367	1	YENK_CAEEL
53	34	47.9	367	1	YENK_CAEEL
54	34	47.9	367	1	YENK_CAEEL
55	34	47.9	367	1	YENK_CAEEL
56	34	47.9	367	1	YENK_CAEEL
57	34	47.9	367	1	YENK_CAEEL
58	34	47.9	367	1	YENK_CAEEL
59	34	47.9	367	1	YENK_CAEEL
60	34	47.9	367	1	YENK_CAEEL
61	34	47.9	367	1	YENK_CAEEL
62	34	47.9	367	1	YENK_CAEEL
63	34	47.9	367	1	YENK_CAEEL
64	34	47.9	367	1	YENK_CAEEL
65	34	47.9	367	1	YENK_CAEEL
66	34	47.9	367	1	YENK_CAEEL
67	34	47.9	367	1	YENK_CAEEL
68	34	47.9	367	1	YENK_CAEEL
69	34	47.9	367	1	YENK_CAEEL
70	34	47.9	367	1	YENK_CAEEL
71	34	47.9	367	1	YENK_CAEEL
72	34	47.9	367	1	YENK_CAEEL
73	34	47.9	367	1	YENK_CAEEL
74	34	47.9	367	1	YENK_CAEEL
75	34	47.9	367	1	YENK_CAEEL
76	34	47.9	367	1	YENK_CAEEL
77	34	47.9	367	1	YENK_CAEEL
78	34	47.9	367	1	YENK_CAEEL
79	34	47.9	367	1	YENK_CAEEL
80	34	47.9	367	1	YENK_CAEEL
81	34	47.9	367	1	YENK_CAEEL
82	34	47.9	367	1	YENK_CAEEL
83	34	47.9	367	1	YENK_CAEEL
84	34	47.9	367	1	YENK_CAEEL
85	34	47.9	367	1	YENK_CAEEL
86	34	47.9	367	1	YENK_CAEEL
87	34	47.9	367	1	YENK_CAEEL
88	34	47.9	367	1	YENK_CAEEL
89	34	47.9	367	1	YENK_CAEEL
90	34	47.9	367	1	YENK_CAEEL
91	34	47.9	367	1	YENK_CAEEL
92	34	47.9	367	1	YENK_CAEEL
93	34	47.9	367	1	YENK_CAEEL
94	34	47.9	367	1	YENK_CAEEL
95	34	47.9	367	1	YENK_CAEEL
96	34	47.9	367	1	YENK_CAEEL
97	34	47.9	367	1	YENK_CAEEL
98	34	47.9	367	1	YENK_CAEEL
99	34	47.9	367	1	YENK_CAEEL
100	34	47.9	367	1	YENK_CAEEL

FT CARBOHYD 180 180 POTENTIAL.  
 FT CARBOHYD 459 459 POTENTIAL.  
 FT CARBOHYD 545 545 POTENTIAL.  
 SQ SEQUENCE 999 AA; 107503 MW; 4891F6AE CRC32;

Query Match 100.0%; Score 71; DB 1; Length 999;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-05;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 LNSKIARFVSOEPA 15  
 |||||||  
 Db 190 LNSKIARFVSOEPA 204

RESULT 2  
 DSG1\_BOVIN STANDARD; PRT; 1043 AA.

AC 003763;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
 DSG1.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;  
 OC Bovinae; Bos.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUZZLE EPITHELIUM;  
 RA KOCH P.J., GOLDSCHMIDT M.D., ZIMBELMANN R., FRANK W.W.;  
 RN Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE OF 44-1043 FROM N.A.  
 RC TISSUE=MUZZLE EPITHELIUM;  
 RX MEDLINE: 91168965.  
 RA KOCH P.J., WALSH M.J., SCHMELZ M., GOLDSCHMIDT M.D.,  
 RA ZIMBELMANN R., FRANK W.W.;  
 RT "Identification of desmoglein, a constitutive desmosomal  
 glycoprotein, as a member of the cadherin family of cell adhesion  
 molecules."  
 RL Eur. J. Cell Biol. 53:1-12(1990).  
 RN [3]  
 RP REVISIONS, AND SEQUENCE OF 101-123.  
 RX MEDLINE: 92037656.  
 RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMBELMANN R.,  
 RA FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT polypeptide and identification of a second type of desmoglein gene."  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 RN [4]  
 RP SEQUENCE OF 44-493 FROM N.A.  
 RX MEDLINE: 91097553.  
 RA GOODWIN L., HILL J.E., RAYNOR K., RASZI L., NABANE M., COWIN P.;  
 RT "Desmoglein shows extensive homology to the cadherin family of cell  
 adhesion molecules."  
 RL Biochem. Biophys. Res. Commun. 173:1224-1230(1990).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, MUZZLE, TONGUE AND ESOPHAGUS.  
 CC -1- DOMAIN: CALCITONIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY, DESMOsome SUBFAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>)

CC or send an email to [license@isb.ch](mailto:license@isb.ch).  
 CC -----  
 CC EMBL: X58466; CAA11380.1;  
 CC EMBL: X57784; CAA40930.1;  
 CC EMBL: M58165; AAA62709.1;  
 CC PIR: S14603; IYBOG1.  
 CC HSP: P09803; 1EDH.  
 CC PIR: P00028; cadherin.3.  
 CC PROSITE: PS00232; CADHERIN; 2.  
 CC Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 CC Calcium-binding; Repeat.  
 FT SIGNAL 1 23  
 FT PROPEP 24 49  
 FT CHAIN 50 1043  
 FT DOMAIN 50 548  
 FT TRANSMEM 549 573  
 FT DOMAIN 574 1043  
 FT REPEAT 159 158  
 FT REPEAT 159 270  
 FT REPEAT 271 385  
 FT REPEAT 386 498  
 FT REPEAT 499 845  
 FT REPEAT 846 875  
 FT REPEAT 876 905  
 FT REPEAT 906 933  
 FT REPEAT 934 962  
 FT DOMAIN 963 1012  
 FT CARBOHYD 110 110  
 FT CARBOHYD 180 180  
 FT CARBOHYD 496 496  
 FT CONFLICT 124 124  
 SQ SEQUENCE 1043 AA; 112243 MW; 13898584 CRC32;

Query Match 87.3%; Score 62; DB 1; Length 1043;  
 Best Local Similarity 80.0%; Pred. No. 0.00067;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

1 LNSKIARFVSOEPA 15  
 |||||||  
 Db 190 LNSKIARFVSOEPA 204

RESULT 3  
 DSG1\_HUMAN STANDARD; PRT; 1049 AA.

AC 002413;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 1 PRECURSOR (DESMOSOMAL GLYCOPROTEIN 1) (DGL).  
 DSG1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=KERATINOCYTES;  
 RX MEDLINE: 91271279.  
 RA WHEELER G.N., PARKER A.E., THOMAS C.L., ATALIOIS P., POYNTER D.,  
 RA ARNEJAN J., RUTMAN A.J., PIDSLEY S.C., WATT F.M., REES D.A.,  
 RA BUXTON R.S., MAGEE A.I.;  
 RT "Desmosomal glycoprotein DGL, a component of intercellular desmosome  
 RT junctions, is related to the cadherin family of cell adhesion  
 RT molecules."  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:4796-4800(1991).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOsome JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EPIDERMIS, TONGUE, TONSIL AND ESOPHAGUS.  
 CC -1- DOMAIN: CALCITONIN MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X56654; CAA39976.1; .  
 DR PIR: S16906; IYHUG1.  
 DR HSSP: P09803; 1EDH.  
 DR MIM: 125670; .  
 DR PFAM: PF00028; cadherin; 4.  
 DR PROSITE: PS00232; CADHERIN; 2.  
 KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein;  
 KM Calcium-binding; Repeat.  
 FT SIGNAL 1 23  
 FT PROPEP 24 49  
 FT CHAIN 50 1049  
 FT DOMAIN 50 545  
 FT TRANSMEM 546 570  
 FT DOMAIN 571 1049  
 FT REPEAT 50 158  
 FT REPEAT 159 270  
 FT REPEAT 271 385  
 FT REPEAT 386 497  
 FT REPEAT 497 839  
 FT REPEAT 840 869  
 FT REPEAT 870 899  
 FT REPEAT 900 927  
 FT REPEAT 928 956  
 FT DOMAIN 969 1019  
 FT CARBOHYD 36 36  
 FT CARBOHYD 110 110  
 FT CARBOHYD 180 180  
 FT SEQUENCE 1049 AA; 113715 MW; FDD79961 CRC32;  
 SQ  
 Query Match 87.3%; Score 62; DB 1; Length 1049;  
 Best Local Similarity 80.0%; Pred. No. 0.00067;  
 Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LNSKIARKIVSQEPA 15  
 DB 190 LNSKIARKIVROEPS 204  
 RESULT 4  
 ID DS62\_HUMAN STANDARD; PRT; 1117 AA.  
 AC Q14126;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE DESMOGLEIN 2 PRECURSOR (HDGC).  
 GN DS62.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-COLON CARCINOMA;  
 RX SCHAEFER S., KOCH P.J., FRANK W.W.;  
 RA "Identification of the ubiquitous human desmoglein, Dsg2, and the  
 RT expression catalogue of the desmoglein subfamily of desmosomal  
 RT cadherins";  
 RL Exp. Cell Res. 211:391-399(1994).  
 RN [2]  
 RP SEQUENCE OF 777-1117 FROM N.A.  
 RX MEDLINE; 92037656.

RA KOCH P.J., GOLDSCHMIDT M.D., WALSH M.J., ZIMMELMANN R., FRANK W.W.;  
 RT "Complete amino acid sequence of the epidermal desmoglein precursor  
 RT polypeptide and identification of a second type of desmoglein gene";  
 RL Eur. J. Cell Biol. 55:200-208(1991).  
 CC -1- FUNCTION: COMPONENT OF INTERCELLULAR DESMOSOME JUNCTIONS.  
 CC INVOLVED IN THE INTERACTION OF PLAQUE PROTEINS AND INTERMEDIATE  
 CC FILAMENTS MEDIATING CELL-CELL ADHESION.  
 CC SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 CC -1- TISSUE SPECIFICITY: ALL OF THE TISSUES TESTED AND CARCINOMAS.  
 CC -1- DOMAIN: CALCIUM MAY BE BOUND BY THE CADHERIN-LIKE REPEATS  
 CC (POTENTIAL).  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY. DESMOSOMAL SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z26317; CAA81226.1; .  
 DR HSSP: P15116; INCI.  
 DR MIM: 125671; .  
 DR PFAM: PF00028; cadherin; 4.  
 DR PROSITE: PS00232; CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Repeat; Signal;  
 KM Cytoskeleton; Calcium-binding.  
 FT SIGNAL 1 23  
 FT PROPEP 24 48  
 FT CHAIN 49 1117  
 FT DOMAIN 49 608  
 FT TRANSMEM 609 633  
 FT DOMAIN 634 1117  
 FT REPEAT 49 159  
 FT REPEAT 160 272  
 FT REPEAT 273 387  
 FT REPEAT 388 502  
 FT REPEAT 502 911  
 FT REPEAT 912 941  
 FT REPEAT 942 967  
 FT REPEAT 968 991  
 FT REPEAT 992 1020  
 FT REPEAT 1021 1050  
 FT CARBOHYD 111 111  
 FT CARBOHYD 181 181  
 FT CARBOHYD 308 308  
 FT CARBOHYD 461 461  
 FT CARBOHYD 513 513  
 FT SEQUENCE 1117 AA; 122385 MW; 84D3B898 CRC32;  
 SQ  
 Query Match 77.5%; Score 55; DB 1; Length 1117;  
 Best Local Similarity 73.3%; Pred. No. 0.015;  
 Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 LNSKIARKIVSQEPA 15  
 DB 191 LNSKIARKIVSLEPA 205  
 RESULT 5  
 ID CAD1\_MOUSE STANDARD; PRT; 884 AA.  
 AC P09803; Q61377;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (VIMORULIN) (ARC-1).  
 GN CDH1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-1CR.  
 RX MEDLINE: 87315445.  
 RA NAGAFUCHI A., SHIRAYOSHI Y., OKAZARI K., YASUDA K., TAKEICHI M.;  
 RT "Transformation of cell adhesion properties by exogenously introduced  
 E-cadherin cDNA";  
 RL Nature 329:341-343(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-129/SV.  
 RX MEDLINE: 92093614.  
 RA RINGMANN M., BARIBAUPT H., SCHMIDT C., KEMLER R.;  
 RT "The structure of the gene coding for the mouse cell adhesion  
 molecule uvomorulin";  
 RL Nucleic Acids Res. 19:6533-6539(1991).  
 RN [3]  
 RP SEQUENCE OF 174-884 FROM N.A., AND SEQUENCE OF 157-181.  
 RX MEDLINE: 88111553.  
 RA RINGMANN M., SCHUB R., WESTMEIER D., EISENBERG H., LOTTSPEICH F.,  
 ENGEL J., DOELZ R., JAEHNIG F., EPPLEN J., MAYER S., MUELLER C.,  
 KEMLER R.;  
 RT "The structure of cell adhesion molecule uvomorulin. Insights into  
 the molecular mechanism of Ca<sup>2+</sup>-dependent cell adhesion";  
 RL EMBO J. 6:3647-3653(1987).  
 RN [4]  
 RP SEQUENCE OF 1-15 FROM N.A.  
 RX MEDLINE: 92107977.  
 RA BEHRENS J., LOEWERICK O., KLEIN-HITPASS L., BIRCHMEIER W.;  
 RT "The E-cadherin promoter: functional analysis of a G-C-rich region  
 and an epithelial cell-specific palindromic regulatory element";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11495-11499(1991).  
 RN [5]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 157-370.  
 RX MEDLINE: 96176249.  
 RA NAGAR B., OVERDUIN M., IKURA M., RINI J.M.;  
 RT "Structural basis of calcium-induced E-cadherin rigidification and  
 dimerization";  
 RL Nature 380:360-364(1996).  
 RN [6]  
 RP STRUCTURE BY NMR OF 157-260.  
 RX MEDLINE: 96271285.  
 RA OVERDUIN M., TONG K.I., KAY C.M., IKURA M.;  
 RT "1H, 15N and 13C resonance assignments and monomeric structure of the  
 amino-terminal extracellular domain of epithelial cadherin";  
 RL J. Biomol. NMR 7:173-189(1996).  
 CC - FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC - SUBUNIT: HOMODIMER.  
 CC - TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.  
 CC - SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X60970: CAA43292.1; JOINED.  
 DR EMBL: X60971: CAA43292.1; JOINED.  
 DR EMBL: X60972: CAA43292.1; JOINED.  
 DR EMBL: X60973: CAA43292.1; JOINED.  
 DR EMBL: X60974: CAA43292.1; JOINED.  
 DR EMBL: X60975: CAA43292.1; JOINED.  
 DR EMBL: X60976: CAA43292.1; JOINED.  
 DR EMBL: X60967: CAA43292.1; JOINED.  
 DR EMBL: X60968: CAA43292.1; JOINED.  
 DR EMBL: X60969: CAA43292.1; JOINED.

DR EMBL: X60970: CAA43292.1; JOINED.  
 DR EMBL: X60971: CAA43292.1; JOINED.  
 DR EMBL: X60972: CAA43292.1; JOINED.  
 DR EMBL: X60973: CAA43292.1; JOINED.  
 DR EMBL: X60974: CAA43292.1; JOINED.  
 DR EMBL: X60975: CAA43292.1; JOINED.  
 DR EMBL: X60976: CAA43292.1; JOINED.  
 DR EMBL: X60967: CAA43292.1; JOINED.  
 DR EMBL: X60968: CAA43292.1; JOINED.  
 DR EMBL: X60969: CAA43292.1; JOINED.  
 DR PIR: S04528; IUMSC.  
 DR PDB: 1EDH; 11-JUN-97.  
 DR PDB: 1SUH; 11-JUL-96.  
 DR MGD: MGI:88354; CDH1.  
 DR PFAM: PF00028; cadherin. 5.  
 DR PFAM: PF01049; cadherin.C term. 1.  
 DR PROSITE: PS00232; CADHERIN. 3.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KM Calcium-binding; Repeat; Signal; 3D-structure.  
 FT SIGNAL 1 27  
 FT PROPEP 28 156  
 FT CHAIN 157 884  
 FT DOMAIN 157 709  
 FT TRANSMEM 710 733  
 FT DOMAIN 734 884  
 FT REPEAT 157 264  
 FT REPEAT 265 377  
 FT REPEAT 378 488  
 FT REPEAT 489 595  
 FT REPEAT 596 699  
 FT DOMAIN 840 855  
 FT CARBOHYD 560 560  
 FT CARBOHYD 639 639  
 FT CONFLICT 267 267  
 FT CONFLICT 272 272  
 FT SEQUENCE 884 AA: 98255 MM; 86758910 CRC32;  
 Query Match 59.2%; Score 42; DB 1; Length 884;  
 Best Local Similarity 61.5%; Pred. No. 3.6;  
 Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 NSKIATVTSQDP 14  
 Db 299 NAAIATVTSQDP 311  
 RESULT 6  
 CDB\_CHICK STANDARD; PRT; 732 AA.  
 AC P33145;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)  
 DE B-CADHERIN PRECURSOR (K-CAM PROTEIN) (FRAGMENT).  
 GN Gallus gallus (Chicken).  
 OS Neophala; Metazoa; Chordata; Vertebrata; Archosauria; Aves;  
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;  
 CC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92107987.  
 RA SORIKIN B.C., GALLIN W.J., EDELMAN G.M., CUNNINGHAM B.A.;  
 RT "Genes for two calcium-dependent cell adhesion molecules have similar  
 structures and are arranged in tandem in the chicken genome";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:11545-11549(1991).  
 RN [2]  
 RP SEQUENCE OF 7-733 FROM N.A.  
 RC TISSUE-EMBRYONIC BRAIN;  
 RX MEDLINE: 91225083.  
 RA NAOLITANO E.W., VENSTROM K., WHEELER E.F., REICHARDT L.F.;  
 RT "Molecular cloning and characterization of B-cadherin, a novel chick  
 cadherin";  
 RL J. Cell Biol. 113:893-905(1991).  
 CC - FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.

THEY PRESENTLY INTERACT WITH THEMSELVES IN A HOMOPHILIC MANNER IN CONNECTING CELLS. CADHERINS MAY THIS CONTRIBUTE TO THE SORTING OF HETEROGENEOUS CELL TYPES. B-CADHERIN MAY HAVE IMPORTANT FUNCTIONS IN NEUROGENESIS, IN AT LEAST SOME EPITHELIA, AND IN EMBRYOGENESIS.

-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

-1- TISSUE SPECIFICITY: EXPRESSED IN A WIDE VARIETY OF TISSUES.

-1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: M81894; AAA48929.1; -

DR EMBL: X58518; CAA41408.1; -

DR PIR: A41634; IYCHCB.

DR HSSP: P09803; ISUH.

DR PFAM: PF00028; cadherin; 5.

DR PFAM: PF01049; cadherin\_C-term; 1.

DR PROSITE: PS00232; CADHERIN; 3.

KW Cell adhesion; glycoprotein; phosphorylation; Transmembrane; Calcium-binding; Repeat.

KW NON\_TER 1

FT PROPEP <1 6 POTENTIAL.

FT CHAIN 7 732 B-CADHERIN.

FT DOMAIN 6 554 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 555 580 POTENTIAL.

FT DOMAIN 581 732 CYTOPLASMIC (POTENTIAL).

FT REPEAT 6 114 CADHERIN 1.

FT REPEAT 115 227 CADHERIN 2.

FT REPEAT 228 339 CADHERIN 3.

FT REPEAT 340 443 CADHERIN 4.

FT REPEAT 444 554 CADHERIN 5.

FT DOMAIN 689 702 SER-RICH.

FT CARBOHYD 137 137 POTENTIAL.

FT CARBOHYD 410 410 POTENTIAL.

FT CONFLICT 414 414 M -> V (IN REF. 2).

SO SEQUENCE 732 AA; 80613 MW; 9C3CF5E9 CRC32.

Query Match 57.78; Score 41; DB 1; Length 732;  
Best Local Similarity 61.58; Pred. No. 4.5;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 NSKIAFKIVSOEP 14  
| 11:111111|  
Db 149 NGVIAYSILSOEP 161

RESULT 7  
FAT\_DROME STANDARD; PRT; 5147 AA.

AC P33450;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE CADHERIN-RELATED TUMOR SUPPRESSOR PRECURSOR (FAT PROTEIN).  
GN FT.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 92069752.  
RA MAHONEY P.A., WEBER U., ONOFRECHUK P., BISSMANN H., BRYANT P.J.,  
RT "The fat tumor suppressor gene in Drosophila encodes a novel member  
of the cadherin gene superfamily.";

Cell 67:853-868(1991).

CC -1- FUNCTION: COULD FUNCTION AS A CELL-ADHESION PROTEIN.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- DISEASE: RECESSIVE LETHAL MUTATIONS IN FAT CAUSE HYPERPLASTIC, TUMOR-LIKE OVERGROWTH OF LARVAL IMAGINAL DISCS, DEFECTS IN DIFFERENTIATION AND MORPHOGENESIS, AND DEATH DURING THE PUPAL STAGE.

CC -1- MISCELLANEOUS: SIMILARITY: CONTAINS 37 CADHERIN-TYPE REPEATS.

CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.

CC -1- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----

DR EMBL: M80537; AA28530.1; -

DR PIR: A41087; IYFTM.

DR HSSP: P00740; IIXA.

DR FLYBASE: FBgn001075; ft.

DR PFAM: PF00008; EGF; 4.

DR PFAM: PF00028; cadherin; 34.

DR PFAM: PF00054; laminin\_G; 2.

DR PROSITE: PS00232; CADHERIN; 22.

DR PROSITE: PS01066; EGF\_2; 2.

KW Cell adhesion; Signal; Transmembrane; Cytoskeleton; Glycoprotein; Calcium-binding; Repeat; EGF-like domain.

KW CHAIN 1 35 POTENTIAL.

FT DOMAIN 36 5147 CADHERIN-RELATED TUMOR SUPPRESSOR.

FT TRANSMEM 4584 4609 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 4610 5147 POTENTIAL.

FT REPEAT 35 156 CYTOPLASMIC (POTENTIAL).

FT REPEAT 157 270 CADHERIN 1.

FT REPEAT 271 382 CADHERIN 2.

FT REPEAT 383 494 CADHERIN 3.

FT REPEAT 495 599 CADHERIN 4.

FT REPEAT 600 708 CADHERIN 5.

FT REPEAT 709 820 CADHERIN 6.

FT REPEAT 821 942 CADHERIN 7.

FT REPEAT 943 1049 CADHERIN 8.

FT REPEAT 1050 1153 CADHERIN 9.

FT REPEAT 1154 1278 CADHERIN 10.

FT REPEAT 1279 1384 CADHERIN 11.

FT REPEAT 1385 1489 CADHERIN 12.

FT REPEAT 1490 1601 CADHERIN 13.

FT REPEAT 1602 1713 CADHERIN 14.

FT REPEAT 1714 1823 CADHERIN 15.

FT REPEAT 1824 1922 CADHERIN 16.

FT REPEAT 1923 2027 CADHERIN 17.

FT REPEAT 2028 2167 CADHERIN 18.

FT REPEAT 2168 2278 CADHERIN 19.

FT REPEAT 2279 2385 CADHERIN 20.

FT REPEAT 2386 2491 CADHERIN 21.

FT REPEAT 2492 2596 CADHERIN 22.

FT REPEAT 2597 2703 CADHERIN 23.

FT REPEAT 2704 2810 CADHERIN 24.

FT REPEAT 2811 2913 CADHERIN 25.

FT REPEAT 2914 3013 CADHERIN 26.

FT REPEAT 3014 3124 CADHERIN 27.

FT REPEAT 3125 3229 CADHERIN 28.

FT REPEAT 3230 3334 CADHERIN 29.

FT REPEAT 3335 3439 CADHERIN 30.

FT REPEAT 3440 3545 CADHERIN 31.

FT REPEAT 3546 3651 CADHERIN 32.

FT REPEAT 3652 3756 CADHERIN 33.

FT REPEAT 3757 3852 CADHERIN 34.

FT DOMAIN 4013 4049 EGF-LIKE 1.

FT DOMAIN 4013 4049 EGF-LIKE 2.

FT DOMAIN 4052 4090 EGF-LIKE 3.  
 FT DOMAIN 4092 4128 EGF-LIKE 4.  
 FT DOMAIN 4321 4362 EGF-LIKE 5.  
 FT DISULFID 3954 3966 BY SIMILARITY.  
 FT DISULFID 3960 3999 BY SIMILARITY.  
 FT DISULFID 4001 4010 BY SIMILARITY.  
 FT DISULFID 4017 4028 BY SIMILARITY.  
 FT DISULFID 4037 4057 BY SIMILARITY.  
 FT DISULFID 4039 4048 BY SIMILARITY.  
 FT DISULFID 4056 4067 BY SIMILARITY.  
 FT DISULFID 4061 4078 BY SIMILARITY.  
 FT DISULFID 4080 4089 BY SIMILARITY.  
 FT DISULFID 4096 4107 BY SIMILARITY.  
 FT DISULFID 4101 4116 BY SIMILARITY.  
 FT DISULFID 4118 4127 BY SIMILARITY.  
 FT DISULFID 4325 4341 BY SIMILARITY.  
 FT DISULFID 4334 4350 BY SIMILARITY.  
 FT DISULFID 4352 4361 BY SIMILARITY.  
 FT CARBOHYD 239 239 POTENTIAL.  
 FT CARBOHYD 257 257 POTENTIAL.  
 FT CARBOHYD 276 276 POTENTIAL.  
 FT CARBOHYD 280 280 POTENTIAL.  
 FT CARBOHYD 402 402 POTENTIAL.  
 FT CARBOHYD 461 461 POTENTIAL.  
 FT CARBOHYD 605 605 POTENTIAL.  
 FT CARBOHYD 631 631 POTENTIAL.  
 FT CARBOHYD 1155 1155 POTENTIAL.  
 FT CARBOHYD 1367 1367 POTENTIAL.  
 FT CARBOHYD 1458 1458 POTENTIAL.  
 FT CARBOHYD 1751 1751 POTENTIAL.  
 FT CARBOHYD 1831 1831 POTENTIAL.  
 FT CARBOHYD 1880 1880 POTENTIAL.  
 FT CARBOHYD 2080 2080 POTENTIAL.  
 FT CARBOHYD 2171 2171 POTENTIAL.  
 FT CARBOHYD 2247 2247 POTENTIAL.  
 FT CARBOHYD 2290 2290 POTENTIAL.  
 FT CARBOHYD 2437 2437 POTENTIAL.  
 FT CARBOHYD 2581 2581 POTENTIAL.  
 FT CARBOHYD 2789 2789 POTENTIAL.  
 FT CARBOHYD 2920 2920 POTENTIAL.  
 FT CARBOHYD 2946 2946 POTENTIAL.  
 FT CARBOHYD 2967 2967 POTENTIAL.  
 FT CARBOHYD 3167 3167 POTENTIAL.  
 FT CARBOHYD 3303 3303 POTENTIAL.  
 FT CARBOHYD 3386 3386 POTENTIAL.  
 FT CARBOHYD 3389 3389 POTENTIAL.  
 FT CARBOHYD 3525 3525 POTENTIAL.  
 FT CARBOHYD 3852 3852 POTENTIAL.  
 FT CARBOHYD 3865 3865 POTENTIAL.  
 FT CARBOHYD 3905 3905 POTENTIAL.  
 FT CARBOHYD 4306 4306 POTENTIAL.  
 FT CARBOHYD 4414 4414 POTENTIAL.  
 FT CARBOHYD 4471 4471 POTENTIAL.  
 FT CARBOHYD 4487 4487 POTENTIAL.  
 FT CARBOHYD 4539 4539 POTENTIAL.  
 FT CARBOHYD 4550 4550 POTENTIAL.  
 FT VARIANT 1229 1229 S->G.  
 FT VARIANT 1233 1233 G->S.  
 SQ SEQUENCE 5147 AA; 564868 MW; 1EF20E13 CRC32;

Query Match 56.3%; Score 40; DB 1; Length 5147;  
 Best Local Similarity 50.0%; Pred No. 54;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEP 14  
 DB 1520 LNKVSYAISKQEP 1533  
 RESULT 8  
 CADI\_HUMAN  
 ID CADI\_HUMAN STANDARD: PRT; 882 AA.

AC P12830; 014216;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE EPITHELIAL-CADHERIN PRECURSOR (E-CADHERIN) (UWOMORULIN) (CAM 120/80).  
 DR CDH1 OR UVO OR CDHE.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OC [1]  
 RP SEQUENCE FROM N.A.  
 RP MEDLINE: 93211394.  
 RA BUSSEMAKERS M.J.G., MEES S.G.M., VAN BOKHOVEN A., DEBRUYNE F.M.J.,  
 RA SCHALKEN J.A.;  
 RT "Molecular cloning and characterization of the human E-cadherin  
 RT cDNA."  
 RL Mol. Biol. Rep. 17:123-128(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA KELLER W., WARD A., ODA T., HIROHASHI S., KEMLER R., BIRCHMEIER W.;  
 RN Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 172-311 FROM N.A.  
 RP TISSUE-LIVER:  
 RC MEDLINE: 89031725.  
 RA MANSOURI A., SPURR N., GOODFELLOW P.N., KEMLER R.;  
 RT "Characterization and chromosomal localization of the gene encoding  
 RT the human cell adhesion molecule uvomorulin."  
 RL Differentiation 38:67-71(1988).  
 RN [4]  
 RP SEQUENCE OF 1-16 FROM N.A.  
 RX MEDLINE: 94380041.  
 RA BUSEMAKERS M.J., GIROLDI L.A., VAN BOKHOVEN A., SCHALKEN J.A.;  
 RT "Transcriptional regulation of the human E-cadherin gene in human  
 RT prostate cancer cell lines: characterization of the human E-cadherin  
 RT gene promoter."  
 RL Biochem. Biophys. Res. Commun. 203:1284-1290(1994).  
 RN [5]  
 RP REVIEW ON VARIANTS.  
 RX MEDLINE: 98415721.  
 RA BERR G., BECKER K.-F., HOEFLE H., VAN ROY F.;  
 RT "Mutations of the human E-cadherin (CDH1) gene."  
 RL Hum. Mutat. 12:226-237(1998).  
 RN [6]  
 RP VARIANTS ALA-370 AND ASN-473.  
 RX MEDLINE: 94306394.  
 RA BECKER K.-F., ATKINSON M.J., REICH U., BECKER I., NEKARDA H.,  
 RA SIEWERT J.R., HOEFLE H.;  
 RT "E-cadherin gene mutations provide clues to diffuse type gastric  
 RT carcinomas."  
 RL Cancer Res. 54:3845-3852(1994).  
 RN [7]  
 RP VARIANT LOBULAR BREAST CARCINOMA SER-315.  
 RX MEDLINE: 95049851.  
 RA KANAI Y., ODA T., TSUDA H., OCHIAI A., HIROHASHI S.;  
 RT "Point mutation of the E-cadherin gene in invasive lobular carcinoma  
 RT of the breast."  
 RL Jpn. J. Cancer Res. 85:1035-1039(1994).  
 RN [8]  
 RP VARIANTS GYNECOLOGIC CANCERS THR-617, VAL-711 AND GLY-838.  
 RX MEDLINE: 94355985.  
 RA RISINGER J.I., BERCHUCK A., KOHLER M.F., BOYD J.;  
 RT "Mutations of the E-cadherin gene in human gynecologic cancers."  
 RL Nat. Genet. 7:98-102(1994).  
 RN [9]  
 RP VARIANT GASTRIC ADENOCARCINOMA GLY-274--PRO-277 DEL.  
 RX MEDLINE: 94173928.  
 RA ODA T., KANAI Y., OYAMA T., YOSHURA K., SHIMOMURA Y., BIRCHMEIER W.,  
 RA SUGIMURA T., HIROHASHI S.;  
 RT "E-cadherin gene-mutations in human gastric carcinoma cell lines."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:1858-1862(1994).  
 RN [10]  
 RP VARIANT GASTRIC CARCINOMA PRO-193.



RX MEDLINE: 96390918.  
 RA MURA H., NOGUCHI M., KANAI Y., OCHIAI A., NAMATA H., HIROHASHI S.;  
 RT "E-cadherin gene mutations in signet ring cell carcinoma of the  
 stomach.";  
 RL Jpn. J. Cancer Res. 87:843-848(1996).  
 RN [11]  
 RP VARIANTS GASTRIC CARCINOMA ASP-400 DEL AND ASP-418--PHE-423 DEL.  
 RX MEDLINE: 97197648.  
 RA TAJURA G., SAKATA K., NISHIZUKA S., MAESAMA C., SUZUKI Y., IWAYA T.,  
 RA TERASHIMA M., SAITO K., SATODATE R.;  
 RT "Inactivation of the E-cadherin gene in primary gastric carcinomas and  
 gastric carcinoma cell lines.";  
 RL Jpn. J. Cancer Res. 87:1153-1159(1996).  
 RN [12]  
 RP VARIANT THYROID TUMOR THR-592.  
 RX MEDLINE: 97138061.  
 RA SOARES P., BEXX G., VAN ROY F., SOBRINHO-SIMÕES M.;  
 RT "E-cadherin gene alterations are rare events in thyroid tumors.";  
 RL Int. J. Cancer 70:32-38(1997).  
 RN [13]  
 RP VARIANTS ASP-336 AND ILE-470.  
 RX MEDLINE: 98196671.  
 RA GUILFORD P., HOPKINS J., HARRAWAY J., MCLEOD M., MCLEOD N.,  
 RA HARAWARA P., TAIRE H., SCOLLAR R., MILLER A., REEVE A.E.;  
 RT "E-cadherin germline mutations in familial gastric cancer.";  
 RL Nature 392:402-405(1998).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. E-CADHERIN HAS A POTENT  
 CC INVASIVE SUPPRESSOR ROLE.  
 CC -1- SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: NON-NEURAL EPITHELIAL TISSUES.  
 CC -1- DISEASE: DEFECTS OF CDH1 IS REGARDED AS ONE OF THE MAIN MOLECULAR  
 CC EVENTS INVOLVED IN DYSFUNCTION OF THE CELL-CELL ADHESION SYSTEM,  
 CC TRIGGERING CANCER INVASION (GASTRIC, BREAST, OVARY, ENDOMETRIUM  
 CC AND THYROID) AND METASTASIS.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z13009; CAA78353.1; -  
 DR EMBL: Z18923; CAA79356.1; -  
 DR EMBL: X12790; CAA31279.1; -  
 DR EMBL: L34545; AAA21764.1; -  
 DR PIR: S25141; IJHUCE.  
 DR PIR: S37654; S37654.  
 DR HSSP: P09803; ISUH.  
 DR MIM: 192090; -  
 DR PFM: PF00028; cadherin; 5.  
 DR PFM: PF01049; Cadherin\_C-term; 1.  
 DR PROSITE: PS00232; Cadherin; 3.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KW Calcium-binding; Repeat; Signal; Disease mutation; Polymorphism.  
 FT SIGNAL 1 27  
 FT CHAIN 28 154  
 FT CHAIN 155 882  
 FT DOMAIN 155 707  
 FT TRASMEN 708 731  
 FT DOMAIN 732 882  
 FT REPEAT 135 262  
 FT REPEAT 263 375  
 FT REPEAT 376 486  
 FT REPEAT 487 593  
 FT REPEAT 594 697  
 FT DOMAIN 838 851  
 FT SER-RICH.

FT CARBOHYD 558 558  
 FT CARBOHYD 637 637  
 FT VARIANT 123 123  
 FT VARIANT 193 193  
 FT VARIANT 274 277  
 FT VARIANT 315 315  
 FT VARIANT 336 336  
 FT VARIANT 370 370  
 FT VARIANT 400 400  
 FT VARIANT 418 423  
 FT VARIANT 463 463  
 FT VARIANT 470 470  
 FT VARIANT 470 470  
 FT VARIANT 470 470  
 FT VARIANT 473 473  
 FT VARIANT 592 592  
 FT VARIANT 598 598  
 FT VARIANT 617 617  
 FT VARIANT 711 711  
 FT VARIANT 838 838  
 FT VARIANT 838 838  
 FT CONFLICT 543 543  
 FT SEQUENCE 882 AA; 97456 MW; 10821AB9 CRC32;  
 SQ  
 Query Match 54.98; Score 39; DB 1; Length 882;  
 Best Local Similarity 53.88; Pred. No. 13;  
 Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 NSKIAFKIVSOEP 14  
 DB 297 NNAIYITLSQDP 309  
 RESULT 9  
 ID CAD2\_HUMAN STANDARD; PRT: 906 AA.  
 AC P19022; Q14923; (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE NEURAL-CADHERIN PRECURSOR (N-CADHERIN).  
 GN CDH2 OR CDHN OR NCAD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91016946.  
 RA REID R.A., HEMPERLY J.J.;  
 RT "Human N-cadherin: nucleotide and deduced amino acid sequence.";  
 RL Nucleic Acids Res. 18:5896-5896(1990).  
 RN [2]  
 RP REVISIONS TO 341, 699 AND 705.  
 RA REID R.A.;

RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 92363956.  
 RA SALOMON D., AYALON O., PATEL-KING R., HYNES R.O., GEIGER B.;  
 RT "Extrafunctional distribution of N-cadherin in cultured human  
 endothelial cells";  
 RL J. Cell Sci. 102:7-17(1992).  
 RN [4]  
 RP SEQUENCE OF 160-906 FROM N.A.  
 RX MEDLINE: 90347462.  
 RA WALSH F.S., BARTON C.H., PUTT W., MOORE S.E., KELSEY D.,  
 RA SPURR N., GOODELLOW P.N.;  
 RT "N-cadherin gene maps to human chromosome 18 and is not linked to the  
 E-cadherin gene";  
 RL J. Neurochem. 55:805-812(1990).  
 RN [5]  
 RP SEQUENCE OF 1-20 FROM N.A.  
 RX MEDLINE: 95048366.  
 RA WALLIS J.A., FOX M., WALSH F.S.;  
 RT "Structure of the human N-cadherin gene: YAC analysis and fine  
 genomic mapping to 18q11.2";  
 RL Genomics 22:172-179(1994).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN  
 CC NEURONAL RECOGNITION MECHANISM.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X57548: CAA0773.1: -  
 DR EMBL: X54315: CAA38213.1: -  
 DR EMBL: S42303: AAB2854.1: -  
 DR EMBL: M34064: AAA03236.1: -  
 DR EMBL: 227420: CAA81799.1: -  
 DR PIR: A38870: ICHUCN.  
 DR HSSP: P15116: INCH.  
 DR WIM: 114020: -  
 DR PFAM: PF00028: cadherin; 5.  
 DR PFAM: PF01049: Cadherin\_C\_term; 1.  
 DR PROSITE: PS00232: CADHERIN; 3.  
 KW Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 KM Calcium-binding; Repeat; Signal.  
 FT SIGNAL 1 23  
 FT PROPEP 24 159  
 FT CHAIN 160 906  
 FT DOMAIN 160 724  
 FT TRANSEM 725 746  
 FT DOMAIN 747 906  
 FT REPEAT 160 267  
 FT REPEAT 268 382  
 FT REPEAT 383 497  
 FT REPEAT 498 603  
 FT REPEAT 604 714  
 FT DOMAIN 863 878  
 FT CARBOHYD 190 190  
 FT CARBOHYD 273 273  
 FT CARBOHYD 325 325  
 FT CARBOHYD 402 402  
 FT CARBOHYD 572 572  
 FT CARBOHYD 651 651  
 FT CARBOHYD 692 692  
 FT CONFLICT 12 12  
 FT CONFLICT 16 16  
 L -> A (IN REF. 3 AND 5).

FT CONFLICT 196 196 S -> T (IN REF. 1).  
 FT CONFLICT 212 212 I -> L (IN REF. 4).  
 FT CONFLICT 357 357 N -> I (IN REF. 1).  
 FT CONFLICT 867 867 A -> L (IN REF. 3).  
 SQ SEQUENCE 906 AA; 99851 MW; BB1F9558 CRC32;  
 Query Match 54.9%; Score 39; DB 1; Length 906;  
 Best Local Similarity 46.7%; Pred. No. 14;  
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 NSKIAKIVSOEPA 15  
 Db 300 LNKMLRYRYSQAPS 314  
 RESULT 10  
 VACB\_MYCPN STANDARD; PRT; 726 AA.  
 ID VACB\_MYCPN  
 AC P75529;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VACB PROTEIN HOMOLOG.  
 GN VACB.  
 OS Mycoplasma pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
 OC Mycoplasmataceae; Mycoplasma.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 29342 / M129;  
 RX MEDLINE: 97105885.  
 RA HILBERT R., HILBERT H., PLAGENS H., PIRKL E., LI B.-C.,  
 RA HERRMANN R.;  
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma  
 pneumoniae";  
 RL Nucleic Acids Res. 24:4420-4449(1996).  
 CC -1- SIMILARITY: BELONGS TO THE RIBONUCLEASE II (RNB) FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AE000057: AAB96237.1: -  
 DR PFAM: PF00575: SL; 1.  
 DR PFAM: PF00773: RNB; 1.  
 DR PROSITE: PS01175: RIBONUCLEASE\_II; 1.  
 KW Hydrolyase; Nuclease.  
 SQ SEQUENCE 726 AA; 83219 MW; C8974499 CRC32;  
 Query Match 54.9%; Score 39; DB 1; Length 726;  
 Best Local Similarity 50.0%; Pred. No. 11;  
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
 QY 2 NSKIAKIVSOEPA 15  
 Db 193 NTKVAFIDEYDPA 206  
 RESULT 11  
 CADB\_XENLA STANDARD; PRT; 905 AA.  
 ID CADB\_XENLA  
 AC P33152;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 38, Last annotation update)  
 DE BLASTOMERE-CADHERIN PRECURSOR (B-CADHERIN).  
 OS Xenopus laevis (African clawed frog).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
 CC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;  
 CC Xenopus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95151580.  
 RA MÜLLER H.A., KÜHL M., FINNEMANN S., SCHNEIDER S., VAN DER POEL S.,  
 RA HAUSEN P., WEDLICH D.;  
 RT "Xenopus cadherins: the maternal pool comprises distinguishable  
 RT members of the family."  
 RL Mech. Dev. 47:213-223(1994).  
 RN [2]  
 RP SEQUENCE OF 459-905 FROM N.A.  
 RX MEDLINE; 92062581.  
 RA HERZBERG F., WILDERMUTH V., WEDLICH D.;  
 RT "Expression of Xcad, a novel cadherin, during oogenesis and early  
 RT development of Xenopus."  
 RL Mech. Dev. 35:33-42(1991).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PITUITARY GLAND, LUNG AND KIDNEY.  
 CC -1- DEVELOPMENTAL STAGE: DURING OOGENESIS AND EARLY DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC  
 DR EMBL; X78546; CA55292.1;  
 DR EMBL; X63719; CA45251.1;  
 DR PIR; S43065; S43065.  
 DR HSSP; P09803; 1SUH.  
 DR PFAM; PF00028; cadherin.5.  
 DR PFAM; PF01049; Cadherin\_C.term.1.  
 DR PROSITE; PS00232; CADHERIN.3.  
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 DR Calcium-binding; Repeat; Signal.  
 KW  
 FT PROPEP 1 178  
 FT CHAIN 179 905  
 FT DOMAIN 179 727  
 FT TRANSMEM 728 751  
 FT DOMAIN 752 905  
 FT DOMAIN 849 873  
 FT DOMAIN 900 905  
 FT CARBOHYD 448 448  
 FT CARBOHYD 581 581  
 FT CARBOHYD 704 704  
 FT CARBOHYD 459 461  
 FT CONFLICT 698 698  
 FT CONFLICT 808 808  
 FT CONFLICT 841 841  
 FT CONFLICT 878 878  
 FT CONFLICT 884 884  
 FT CONFLICT 903 903  
 FT CONFLICT 905 905  
 FT SEQUENCE 905 AA; 100377 MW; 147E1228 CRC32;  
 SQ

Query Match 53.5%; Score 38; DB 1; Length 905;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 1 LNSKIAPKIVSOEP 14  
 11 11:11:11  
 DB 320 LNCVIAISILKODP 333

RESULT 12  
 CADN\_XENLA STANDARD; PRT; 905 AA.  
 ID CADN\_XENLA  
 AC P20310;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE NEURAL-CADHERIN 1 PRECURSOR (N-CADHERIN 1).  
 OS Xenopus laevis (African clawed frog).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;  
 CC Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;  
 CC Xenopus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 90211966.  
 RA DERRICK R.J., DICKER D., KINTNER C.R.;  
 RT "The effects of N-cadherin misexpression on morphogenesis in Xenopus  
 RT embryos."  
 RL Neuron 4:493-506(1990).  
 CC -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.  
 CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC  
 CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE  
 CC SORTING OF HETEROGENEOUS CELL TYPES. N-CADHERIN MAY BE INVOLVED IN  
 CC NEURONAL RECOGNITION MECHANISM.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- SIMILARITY: BELONGS TO THE CADHERIN FAMILY.  
 CC -----  
 CC PIR; J00442; IJYIC1.  
 DR HSSP; P15116; 1MCH.  
 DR PFAM; PF00028; cadherin.5.  
 DR PFAM; PF01048; Cadherin\_C.term.1.  
 DR PROSITE; PS00232; CADHERIN.3.  
 DR Cell adhesion; Glycoprotein; Phosphorylation; Transmembrane;  
 DR Calcium-binding; Repeat; Signal.  
 KW  
 FT SIGNAL 1 28  
 FT PROPEP 29 160  
 FT CHAIN 161 905  
 FT DOMAIN 161 723  
 FT TRANSMEM 724 745  
 FT DOMAIN 746 905  
 FT DOMAIN 746 905  
 FT REPEAT 161 268  
 FT REPEAT 269 383  
 FT REPEAT 384 498  
 FT REPEAT 499 604  
 FT REPEAT 605 713  
 FT REPEAT 713 877  
 FT CARBOHYD 191 191  
 FT CARBOHYD 274 274  
 FT CARBOHYD 326 326  
 FT CARBOHYD 403 403  
 FT CARBOHYD 573 573  
 FT CARBOHYD 623 623  
 FT CARBOHYD 651 651  
 FT CARBOHYD 692 692  
 FT SEQUENCE 905 AA; 100549 MW; 50053CA0 CRC32;  
 SQ

Query Match 53.5%; Score 38; DB 1; Length 905;  
 Best Local Similarity 50.0%; Pred. No. 21;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 NSKIAPKIVSOEPA 15  
 1 1:11:11:11  
 DB 302 NGLRYKILSCPTA 315

RESULT 13  
 CADN\_XENLA STANDARD; PRT; 906 AA.  
 ID CADN\_XENLA  
 AC P33147;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1993 (Rel. 27, Last annotation update)



```

RT "Human gamma-aminobutyraldehyde dehydrogenase (ALDH9): cDNA sequence,
RT genomic organization, polymorphism, chromosomal localization, and
RT tissue expression."
RL Genomics 34:376-380(1996).
RM [2]
RP SEQUENCE OF 32-493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE; 94094820.
RA KURYS G., SHAH P.C., KIKOTOGA A., REED D., AMBROZIAK W.,
RA PIETRUSZKO R.;
RT "Human aldehyde dehydrogenase. cDNA cloning and primary structure of
RT the enzyme that catalyzes dehydrogenation of 4-aminobutyraldehyde."
RL Eur. J. Biochem. 218:311-320(1993).
RN [3]
RP CHARACTERIZATION.
RC TISSUE-LIVER;
RX MEDLINE; 89174735.
RA KURYS G., AMBROZIAK W., PIETRUSZKO R.;
RT "Human aldehyde dehydrogenase. Purification and characterization of a
RT third isozyme with low km for gamma-aminobutyraldehyde."
RL J. Biol. Chem. 264:4715-4721(1989).
CC -1- FUNCTION: HAS A WIDE SUBSTRATE SPECIFICITY AND A LOW KM FOR 4-
CC AMINOBTUTRALDEHYDE. ITS OPTIMAL PH DEPENDS UPON SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: ALDEHYDE + NAD(+) + H(2)O - ACID + NADH.
CC -1- CATALYTIC ACTIVITY: 4-AMINOBTUTANAL + NAD(+) + H(2)O - 4-
CC AMINOBTUTANOL + NADH.
CC -1- SUBUNIT: HOMOTETRAMER.
CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN ADULT LIVER, SKELETAL
CC MUSCLE, AND KIDNEY. LOW LEVELS IN HEART, PANCREAS, LUNG, AND
CC BRAIN.
CC -1- DEVELOPMENTAL STAGE: STRONGLY EXPRESSED IN HUMAN EMBRYONIC BRAIN
CC (GESTATIONAL AGE 12 WEEKS).
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE ALDEHYDE DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U34252; AAB18827.1; -
DR EMBL; X75425; CAAS3176.1; -
DR HSSP; P56533; 1A4S.
DR MIM; 602733; -
DR PFM; PF00171; aldehyd; 1.
DR PROSITE; PS00670; ALDEHYDE_DEHYDR_CYS; 1.
DR PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 231 236 NAD (ADP PART) (BY SIMILARITY).
FT ACT_SITE 253 253 POTENTIAL.
FT ACT_SITE 287 287 POTENTIAL.
FT CONFLICT 115 115 C -> S (IN REF. 2).
FT CONFLICT 149 149 C -> O (IN AA SEQUENCE).
FT CONFLICT 158 158 P -> W (IN AA SEQUENCE).
FT CONFLICT 171 171 A -> R (IN REF. 2).
SQ SEQUENCE 493 AA; 53532 MW; 97C6E858 CRC32;

```

```

Query Match          52.1%; Score 37; DB 1; Length 493;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 2 NSRIAKRIVSOE 13
Db 56 NAKAAKRWISQK 67

```

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 12, 2000, 23:15:19 ; Search time 59.21 Seconds  
(without alignments)  
15.518 Million cell updates/sec

Title: US-08-991-628-3  
Sequence: 1 LNSKIAFKIVSOEPA 15

Scoring table: BLOSUM62

Searched: 199794 seqs, 61255205 residues

Database: SPTREMBL\_11.\*

Word size: 0

Number of hits that pass the threshold: 199794

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	97.2	993	11 035902	035902 mus musculu
2	40	56.3	327	5 093345	093345 caenorhabdi
3	40	56.3	145	6 018926	018926 sus scrofa
4	40	56.3	1421	10 080907	080907 arabidopsis
5	39	54.9	829	4 075309	075309 homo sapien
6	39	54.9	878	4 015855	015855 homo sapien
7	39	54.9	829	6 028634	028634 oryctolagus
8	38	53.5	533	2 054891	054891 streptococ
9	38	53.5	533	2 054989	054989 streptococ
10	38	53.5	138	2 072670	072670 synecocyst
11	38	53.5	967	4 094853	094853 homo sapien
12	38	53.5	273	9 038172	038172 bacterioph
13	37	52.1	236	2 005946	005946 rickettsia
14	37	52.1	461	2 092CY0	092CY0 rickettsia
15	37	52.1	936	4 095206	095206 homo sapien
16	37	52.1	252	12 09YR59	09YR59 melanoplus
17	36	50.7	103	2 053450	053450 mycobacteri
18	36	50.7	475	2 092CY9	092CY9 rickettsia
19	36	50.7	238	11 055075	055075 rickettsia
20	36	50.7	830	11 088338	088338 mus musculu
21	36	50.7	277	11 008558	008558 rattus norv
22	36	50.7	276	11 008559	008559 rattus norv
23	36	50.7	906	11 0921Y3	0921Y3 rattus norv
24	35	49.3	308	2 051057	051057 borrelia bu
25	35	49.3	182	5 025918	025918 plasmidium

26	35	49.3	225	5 020152	020152 caenorhabdi
27	35	49.3	241	8 033766	033766 allomyces m
28	35	49.3	838	9 064330	064330 bacterioph
29	35	49.3	1249	10 023289	023289 arabidopsis
30	35	49.3	2261	10 042793	042793 glycine max
31	35	49.3	775	10 P93205	P93205 lycopersico
32	35	49.3	2228	12 055528	055528 sendai viru
33	35	49.3	2228	12 055530	055530 sendai viru
34	35	49.3	2842	12 036452	036452 hepatitis g
35	35	49.3	1980	12 084185	084185 human para
36	35	49.3	2228	12 098705	098705 sendai viru
37	34	47.9	334	1 058135	058135 pyrococcus
38	34	47.9	380	2 066815	066815 aquifex aeo
39	34	47.9	330	2 087491	087491 staphylococ
40	34	47.9	455	3 014039	014039 schizosacch
41	34	47.9	1272	3 013756	013756 schizosacch
42	34	47.9	464	4 043159	043159 homo sapien
43	34	47.9	591	4 060324	060324 homo sapien
44	34	47.9	716	4 014164	014164 homo sapien
45	34	47.9	813	4 075229	075229 homo sapien

## ALIGNMENTS

RESULT 1  
ID 035902 PRELIMINARY; PRT; 993 AA.  
AC 035902;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
DE DESMOGLEIN 3 (FRAGMENT).  
GN DSG3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
CC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C;  
RA ISHIKAWA H., LI K., UIRTO J.;  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
DR EMBL; U86016; AAB65091.1; -.  
DR PFM; PF00028; cadherin; 4.  
DR PROSITE; PS00232; CADHERIN; 2.  
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
FT NON-TER  
SQ SEQUENCE 993 AA; 107888 MW; 881794BD CRC32;

Query Match 97.2%; Score 69; DB 11; Length 993;  
Best Local Similarity 93.3%; Pred. No. 9.9e-05;  
Matches 14; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 LNSKIAFKIVSOEPA 15  
DB 190 MNSKIAFKIVSOEPA 204  
RESULT 2  
ID 093345 PRELIMINARY; PRT; 327 AA.  
AC 093345;  
DT 01-FEB-1997 (TREMBLrel. 02, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)  
DE C36B1.11 PROTEIN.  
GN C36B1.11  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdillia; Rhabdillidae;  
CC Rhabdillina; Rhabdillidae; Peloderinae; Caenorhabdillia.  
RN (1)

RP SEQUENCE FROM N.A.  
 RA LEMARD N.:  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., COPSET T., COOPER J., COLLSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,  
 RA SMADON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";  
 RT Nature 368:32-38(1994).  
 DR EMBL: 280215; CAB02276.1;  
 SO SEQUENCE 327 AA; 37139 MW; D8A2C259 CRC32;

Query Match 56.3%; Score 40; DB 5; Length 327;  
 Best Local Similarity 80.0%; Pred. No. 9.1;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 4 KIAFKIVSOEP 13  
 Db 81 KIAFKIVSOEP 90

RESULT 3  
 ID 018926 PRELIMINARY; PRT; 145 AA.  
 AC 018926;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE P-CADHERIN (FRAGMENT).  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA LOTZ D.A., ZHENG J.J.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF033826; AAB87087.1;  
 DR PFM: PF00028; cadherin.1.  
 FT NON\_TER 1 1  
 FT NON\_TER 145 145  
 SQ SEQUENCE 145 AA; 15617 MW; 49B7A084 CRC32;

Query Match 56.3%; Score 40; DB 6; Length 145;  
 Best Local Similarity 53.8%; Pred. No. 3.9;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14  
 Db 74 NGVAVYSILSOEP 86

RESULT 4  
 ID 080907 PRELIMINARY; PRT; 1421 AA.  
 AC 080907;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE T19C21.7 PROTEIN.  
 GN T19C21.7.  
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae;  
 OC Arabidopsis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CV. COLUMBIA;  
 RA ROUNSELEY S.D., LIN X., KETCHUM K.A., CROSBY M.L., BRANDON R.C.,  
 RA SYRES S.M., KAUL S., MASON T.M., KERLAVAGE A.R., ADAMS M.D.,  
 RA SOMERVILLE C.R., VENTER J.C.;  
 RT "Arabidopsis thaliana chromosome II BAC T19C21 genomic sequence.";  
 RT Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC004683; AAC28760.1;  
 SO SEQUENCE 1421 AA; 154325 MW; 1B6D2FF1 CRC32;

Query Match 56.3%; Score 40; DB 10; Length 1421;  
 Best Local Similarity 42.9%; Pred. No. 42;  
 Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEP 14  
 Db 262 VDSKISYEITONP 275

RESULT 5  
 ID 075309 PRELIMINARY; PRT; 829 AA.  
 AC 075309;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE KSP-CADHERIN.  
 GN CDH16.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE: 98389630.  
 RA THOMSON R.B., WARD D.C., QUAGGIN S.E., IGARASHI P., MUCKLER Z.E.,  
 RA AROSON P.S.;  
 RT "cDNA cloning and chromosomal localization of the human and mouse  
 RT isoforms of ksp-cadherin.";  
 RL Genomics 51:445-451(1998).  
 CC -I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL: AF016272; AAC34255.1;  
 DR PFM: PF00028; cadherin.6.  
 DR PROSITE: PS00232; CADHERIN.2.  
 KM Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 829 AA; 89923 MW; E44B3DBE CRC32;

Query Match 54.9%; Score 39; DB 4; Length 829;  
 Best Local Similarity 57.1%; Pred. No. 37;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 15  
 Db 160 NSDLRPHILSQAPA 173

RESULT 6  
 ID 015855 PRELIMINARY; PRT; 878 AA.  
 AC 015855; Q16194; Q13799;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE UVOKORULIN PRECURSOR (E-CADHERIN) (ARC-1/UVOKORULIN).  
 GN UVO.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

Query Match          54.9%; Score 39; DB 4; Length 878;
Best Local Similarity 53.8%; Pred. No. 39;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0

OY      2 NSKIAFKIVSQEP 14
      1:|:|:|:|:|
Db      293 NNAIAVTLISQDP 305

RESULT 7
Q28634 Q28634 PRELIMINARY; PRT; 829 AA.
AC Q28634;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE KIDNEY-SPECIFIC CADHERIN PRECURSOR (KSP-CADHERIN).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
   Eutheria; Lagomorpha; Leporidae; Oryctolagus.
   [1]
RN SEQUENCE FROM N.A., AND OF SEQUENCE 570-586.
RP STRAIN-NEW ZEALAND WHITE.
RC MEDLINE; 95340560.
RX THOMSON R.B., IGARASHI P., BEMESDERFER D., KIM R., ABU-ALFA A.,
   SOLEIMANI M., ARONSON P.S.;
   "Isolation and cDNA cloning of Ksp-cadherin, a novel kidney-specific
   member of the cadherin multigene family.";
RL J. Biol. Chem. 270:17594-17601(1995).
RT
RT -1- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: KIDNEY.
CC LIMITED TO THE BASOLATERAL MEMBRANES OF RENAL TUBULAR EPITHELIAL

```

Query Match	54.9%	Score 39;	DB 6;	Length 829;
Best Local Similarity	57.1%	Pred. No. 37;		
Matches	8;	Conservative	2;	Mismatches 4;
				Indels 0;
				Gaps 0;
cy	2	NSKIAKFTVSOCPA	15	
	:	:	:	:
Db	160	NSDLRFHLSQCPA	173	

[illegible]



Query Match	53.5%	Score 38	DB	Length 533
Best Local Similarity	44.4%	Pred. No. 36		
Matches	8	Conservative	5	Mismatches 1
				Gaps 4
				Indels 1

RESULT	10	
P72670		
ID	P72670	PRELIMINARY;
		PRT; 138 AA

```

RESULT 10
P72670 ID P72670 PRELIMINARY; PRT; 138 AA.
AC P72670;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, last annotation update)
DE HYPOTHETICAL 16.6 KD PROTEIN.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RN Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RC MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIUDA M., SASAMOTO S., KIMURA T.,
RA HOSHICHI T., MATSUO A., MURAKI A., NAKAZAKI N., NARDO K., OKUMURA S.,
RA SHIMONO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90899; BAA16572.1; -
KW Hypothetical protein.
SQ SEQUENCE: 138 AA; 16580 MW; CMA36352 CRC32;

```

Query Match      53.5%;    Score 38;    DB 2;    Length 138;

```
QY      3 SKIAFKIVSQEP 14
        :|:|:|:|:|:|
Db      28 AKLAFEVIPQDP 39
```

DR EMBL; AB018296; BAA34473.1; -;  
SQ SEQUENCE 967 AA; 109376 MW; C8749C0A CRC32;

Best Local Similarity 46.28; Pred. No. 67;  
Matches 6; Conservative 4; Mismatches

RESULT	12	
038172		
ID	038172	PRELIMINARY;
038172		PRT; 273 AA

RESULT	12		
038172			
ID	038172	PRELIMINARY:	PRT; 273 AA.
AC	038172;		
DT	01-NOV-1996 (TREMBLrel. 01,		
DT	01-NOV-1996 (TREMBLrel. 01,		
DT	01-NOV-1996 (TREMBLrel. 06,		
DE	MAJOR HEAD PROTEIN.		
GN	MHP.		
OS	Bacteriophage B1.		
OC	Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-TP901-1, TP936-1, C3-11;		
RX	MEDLINE: 96193742.		
RA	JOHNSEN M.G., APPEL K.F., MADSEN P.L., VOGENSEN F.K., HAMMER K.,		
RA	ARNAU J.;		
RT	"A genomic region of lactococcal temperate bacteriophage TP901-1		
RT	encoding major virion proteins."		
RL	Virology 218:306-315(1996).		
DR	EMBL: X64706; CAA59183.1;		
Q	SEQUENCE 273 AA; 28880 MW; E7D3AE03 CRC32;		

Query Match	53.5%	Score 38:	DB 9:	Length 273:
Best Local Similarity	61.5%	Pred. No. 18:		
Matches 8:	Conservative 1:	Mismatches 4:	Indels 0:	Gaps 0:

QY	3 SKIAFKIVSQEPA 15
	I . I I I I I I I

Db 213 SALMFKIVNSPA 225

## RESULT 13

ID 005946 PRELIMINARY; PRT; 236 AA.  
 AC 005946;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PROSPHORBOSYLAMINOIMIDAZOLE-SUCCINOCARBOXAMIDE SYNTHASE.  
 GN PRC OR RP220.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 97419517.  
 RA ANDERSSON J.O., ANDERSSON S.G.E.;  
 RT "Genomic rearrangements during evolution of the obligate  
 intracellular parasite Rickettsia prowazekii as inferred from an  
 RT analysis of 52015 bp nucleotide sequence."  
 RL Microbiology 143:2783-2795(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 99039499.  
 RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,  
 RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,  
 RA ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria."  
 RL Nature 396:133-140(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 99005535.  
 RA ANDERSSON S.G.E.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ235270; CAI14683.1;  
 DR PFM; PF01259; SAICAR\_Synt.1.  
 SQ SEQUENCE 236 AA; 27310 MW; 3123672C CRC32;

Query Match 52.1%; Score 37; DB 2; Length 236;  
 Best Local Similarity 38.5%; Pred. No. 24;  
 Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14  
 Db 207 NEKIGFELIQNEP 219

## RESULT 14

ID 092CY0 PRELIMINARY; PRT; 461 AA.  
 AC 092CY0;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE SODIUM/PANTOTHENATE SYMPORTER (PANF).  
 GN RPS71.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 OC Rickettsiaceae; Rickettsiae; Rickettsia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RX MEDLINE; 99039499.  
 RA ANDERSSON S.G.E., ZOMORODIPOUR A., ANDERSSON J.O.,  
 RA SICHERITZ-PONTEN T., ALSMARK U.C.M., PODOWSKI R.M., NAESLUND A.K.,  
 RA ERIKSSON A.S., WINKLER H.H., KURLAND C.G.;

RT "The Genome Sequence of Rickettsia prowazekii and the Origin of  
 RT Mitochondria."  
 RL Nature 396:133-140(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MADRID E;  
 RA ANDERSSON S.G.E.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ235272; CAI15019.1;  
 SQ SEQUENCE 461 AA; 51674 MW; 4D6121F9 CRC32;

Query Match 52.1%; Score 37; DB 2; Length 461;  
 Best Local Similarity 46.7%; Pred. No. 48;  
 Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 LNSKIAFKIVSOEPA 15  
 Db 276 LNSKIAFKIVNOPS 290

## RESULT 15

ID 095206 PRELIMINARY; PRT; 936 AA.  
 AC 095206;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PROTOCADHERIN.  
 GN PCDH8.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE; 99005535.  
 RA STREHL S., GLATT K., LIU Q.M., GLATT H., LALANDE M.;  
 RT "Characterization of two novel protocadherins (PCDH8 and PCDH9)  
 RT localized on human chromosome 13 and mouse chromosome 14."  
 RL Genomics 53:81-89(1998).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
 DR EMBL; AF061573; AAC70009.1;  
 DR PROSITE; PS00232; CADHERIN; 5.  
 KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.  
 SQ SEQUENCE 936 AA; 98879 MW; 3A6F1EC8 CRC32;

Query Match 52.1%; Score 37; DB 4; Length 936;  
 Best Local Similarity 46.2%; Pred. No. 99;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 NSKIAFKIVSOEP 14  
 Db 646 NGELAFELQOQEP 658

Search completed: January 12, 2000, 23:15:21.  
 Job time: 184 sec

Sat Jan 15 11:45:00 2000

us-08-991-628-3.rspt

Page 6

CC pemphigus vulgaris.  
SQ Sequence 614 AA;

Query Match 100.0%; Score 74; DB 1; Length 614;  
Best Local Similarity 100.0%; Pred. No. 7.8e-06;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATOKITRISGVGID 15  
DB 78 ATOKITRISGVGID 92

RESULT 3  
W04841  
ID W04841 standard; peptide; 15 AA.

AC W04841;  
DE 18-FEB-1997 (first entry)  
KW Self epitope of desmoglein 3, implicated in autoimmune disease.  
KW Tolerisation; self-peptide; antigen; autoimmune disease;  
KW autacanthigen; HLA; human leukocyte antigen; T-cell; thymocyte;  
KW pemphigus vulgaris; desmoglein; multiple sclerosis;  
KW herpes simplex virus; adenovirus; phosphomannomutase;  
KW human papillomavirus; Epstein-Barr virus; DNA polymerase;  
KW influenza; haemagglutinin; reovirus; sigma protein.  
OS Homo sapiens.  
PN W09627387-A1.  
PD 12-SEP-1996.  
PF 07-MAR-1996; U03182.  
PR 07-MAR-1995; US-400796.  
PA (HARD) HARVARD COLLEGE.

PI Strominger JL, Nuchterpfennig KW;  
DR WPI: 96-425216/42.  
PT Pemphigus vulgaris auto-antigens and multiple sclerosis non-self  
PT antigens - useful in disease treatment, and method for  
PT identification of other self and non-self antigens implicated in  
PT auto-immune disease

PS Claim 1; Page 38; 58pp; English.  
CC Pharmaceutical preparations for tolerisation to antigens comprise  
CC either an isolated human non-collagen or non-mysin basic protein  
CC (MBP) polypeptide which is capable of tolerising an individual to an  
CC autoantigen; or an isolated human pathogen polypeptide capable of  
CC tolerising an individual to that polypeptide. In both cases, the  
CC polypeptide (whether self or non-self) includes an amino acid  
CC sequence corresponding to a sequence motif for a MHC class II  
CC protein, such as HLA-DR, which is associated with a human autoimmune  
CC disease and which binds to the polypeptide to activate autoreactive  
CC T-cells in individuals with the autoimmune disease. This peptide is  
CC derived from the human desmoglein 3 protein (amino acids 78-93 (sic))  
CC and is implicated as a self epitope in pemphigus vulgaris. Peptides  
CC derived from the human desmoglein protein are described in W04841-47.  
SQ Sequence 15 AA;

Query Match 100.0%; Score 74; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATOKITRISGVGID 15  
DB 1 ATOKITRISGVGID 15

RESULT 4  
W15489

ID W15489 standard; Protein; 778 AA.

AC W15489;  
DE 17-JUN-1997 (first entry)  
KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
KW Pemphigus foliaceus; autoantibody; constant region; IgG;  
KW extracellular region; antigen; hinge portion; skin;  
KW dermatitis herpetiformis; fusion protein; detection; ss.  
OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers  
FT domain 1..545  
FT /note= "Pemphigus foliaceus antigen protein"

PN J09077800-A.  
PD 25-MAR-1997.  
PF 12-SEP-1995; 260899.  
PR 12-SEP-1995; JP-260899.  
PA (NISH/) NISHIKAWA T.  
DR WPI: 97-241758/22.  
DR P-PSDB; T66428.

PT Pemphigus foliaceus antigen-IgG constant region fusion protein - linked  
PT through the hinge region used to treat pemphigus foliaceus  
PS Claim 1; Page 10-12; 17pp; Japanese.  
CC This sequence represents a fused protein recognised by pemphigus  
CC foliaceus patient autoantibody which comprises the constant region  
CC of IgG linked to the extracellular region of pemphigus foliaceus  
CC antigen protein through the hinge portion. Pemphigus foliaceus is  
CC a chronic, generalised, vesicular and scaling skin eruption similar  
CC to dermatitis herpetiformis. The pemphigus foliaceus antigen fusion  
CC protein is useful to treat pemphigus foliaceus. The antigen is  
CC especially administered through an adsorbent upon which the fusion  
CC protein is immobilised via a carrier. The fusion protein is also  
CC useful for detecting pemphigus foliaceus antibodies which is useful  
CC in immunodiagnosis. The fusion protein has little or no side effects.  
SQ Sequence 778 AA;

Query Match 86.5%; Score 64; DB 1; Length 778;  
Best Local Similarity 80.0%; Pred. No. 0.00073;  
Matches 12; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATOKITRISGVGID 15  
DB 79 ANQOVTVRISGVGID 93

RESULT 5

ID R71376 standard; Protein; 551 AA.

AC R71376;  
DE 21-NOV-1995 (first entry)  
KW Human cystathionine beta-synthase.  
KW Cystathionine; beta synthase; human; homocystinuria.  
OS Homo sapiens.  
PN W09507714-A.  
PD 23-MAR-1995.  
PF 12-SEP-1994; U10203.  
PR 13-SEP-1993; US-120960.  
PA (COLS ) UNITV COLORADO.  
PI Kraus JP;  
DR WPI: 95-131185/17.  
DR N-PSDB; O87430.

PT Purified DNA encoding human cystathionine beta-synthase - useful  
PT for producing human cystathionine beta-synthase, used for  
PT treating homocystinuria  
PS Claim 7; Fig 1a-c; 45pp; English.  
CC The amino acid sequence shown in R71376 is the human cystathionine  
CC beta-synthase (CBS). Human CBS can be used to treat patients with  
CC homocystinuria and the cDNA sequence, O87430, from which it is  
CC derived is also useful for screening CBS deficient patients for  
CC mutations in the CBS gene.  
SQ Sequence 551 AA;

Query Match 56.8%; Score 42; DB 1; Length 551;  
Best Local Similarity 42.9%; Pred. No. 6;  
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 2 TOKITRISGVGID 15  
DB 296 TEOTTYEVEGIGID 309

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 6, 2000, 15:12:33 : Search time 51.33 seconds

(without alignments)  
6.922 Million cell updates/sec

Title: US-08-991-628-1

Perfect score: 74

Sequence: 1 ATQKITYRISGVGID 15

Scoring table: BLOSUM62

Searched: 188963 seqs, 2366106 residues

Database: A\_Geneseq\_36.\*

Word size: 0

Number of hits that pass the threshold: 188963

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	74	100.0	999	R30742	Human pemphigus vu
2	74	100.0	614	W07908	Pemphigus vulgaris
3	74	100.0	15	W04841	Self epitope of de
4	64	86.5	778	W15489	Pemphigus foliaceu
5	42	56.8	551	R71376	Human cystathionin
6	39	52.7	504	W60670	Human glaucoma ass
7	39	52.7	504	W64669	Human TIGR protein
8	39	52.7	504	W70496	Triabecular meshwor
9	39	52.7	504	W73500	Human trabecular m
10	39	52.7	504	W89391	MG-6 antigen. Diag
11	37	50.0	309	R06441	Segment of desmoso
12	37	50.0	1363	R11061	Bovine Coronavirus
13	37	50.0	560	W13009	Segment of desmoso
14	37	50.0	1363	W31706	Bovine Coronavirus
15	37	50.0	390	Y04860	Mycobacterium spec
16	36	48.6	712	R27823	Sequence encoded b
17	36	48.6	717	R27824	Sequence encoded b
18	36	48.6	1065	W33819	Arbidopsis celluli
19	36	48.6	289	Y06965	E. Canis P30-6 pro
20	35	47.3	1560	P94145	S. ciemoris protei
21	35	47.3	1962	R10560	Mutant protease (K
22	35	47.3	1962	R10561	Mutant protease (N
23	35	47.3	1962	R10557	Mutant protease (A
24	35	47.3	1962	R10558	Mutant protease (A
25	35	47.3	1962	R10559	Mutant protease (A
26	35	47.3	1959	R10562	Mutant protease (d
27	35	47.3	1964	R10563	Mutant protease (K
28	35	47.3	1974	R10940	Mutant protease (d
29	35	47.3	1968	R10941	Mutant protease (d
30	35	47.3	985	R42995	Glycosyltransferas
31	35	47.3	985	R42214	Aspergillus niger
32	35	47.3	878	R5060	Sequence of human
33	35	47.3	878	R5487	Human E-cadherin p
34	35	47.3	916	W13129	Full length human
35	35	47.3	916	W25658	Human cadherin-4.
36	35	47.3	626	W37140	Pes plastridial pho
37	34.5	46.6	593	R86292	Pseudomonas mesoc
38	34	45.9	304	P70637	Sequence encoded b
39	34	45.9	589	P60303	Sequence encoded b

40	34	45.9	896	1	R63533	Human HT-1376 cell
41	34	45.9	23	1	R65100	Random biotinylati
42	34	45.9	23	1	W46703	Biotinylation pept
43	34	45.9	626	1	W55115	Streptococcus pneu
44	33	44.6	100	1	R37660	Pseudomonas 7a glu
45	33	44.6	336	1	R59739	Pseudomonas glutam

## ALIGNMENTS

RESULT 1  
R30742 standard; Protein: 999 AA.  
ID R30742;  
AC R30742;  
DE 14-JUN-1993 (first entry)  
DE Human pemphigus vulgaris 130KD antigen.  
KW Pemphigus vulgaris; skin disease; autoantibodies;  
KW Keratinocyte cell surface antigen; glycoprotein; cell adhesion.  
OS Homo sapiens.  
PN US7798918.A.  
PD 15-DEC-1992.  
PE 27-NOV-1991; 798918.  
PR 27-NOV-1991; US-798918.  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.  
PI Amagai M, Klaus-Kovtun V, Stanley JR;  
DR WPI 93-06/436/08.  
DR N-PsDB: Q35992.  
PT DNA encoding pemphigus vulgaris antigen - useful in proteins for  
PT diagnostic and therapeutic uses  
PS Disclosure: fig 7; 50pp; English.  
CC This sequence is the pemphigus vulgaris 130KD antigen. The protein  
CC and its encoding DNA may be used in the diagnosis and treatment of  
CC pemphigus vulgaris. It is thought that the antigen may be a cell  
CC adhesion molecule.  
SQ Sequence 999 AA:

Query Match 100.0%; Score 74; DB 1; Length 999;  
Best Local Similarity 100.0%; Pred. No. 1.3e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ATQKITYRISGVGID 15  
DB 79 ATQKITYRISGVGID 93

RESULT 2  
W07908 standard; protein: 614 AA.  
ID W07908  
AC W07908; (first entry)  
DE Pemphigus vulgaris antigen protein extracellular region.  
DE Autoantibody; immunoglobulin G; IgG1; fusion protein; diagnosis;  
KW treatment; pemphigus vulgaris; PV; bulla; blister; skin disease;  
KW dermatology.  
OS Homo sapiens.  
PN J08188540-A.  
PD 23-JUL-1996.  
PE 30-JUN-1995; 165632.  
PR 30-JUN-1994; JP-173291.  
PA (NISH/) NISHIKAWA T.  
DR WPI: 96-38852/39.  
PT Fused protein recognised by pemphigus vulgaris autoantibody -  
PT useful to treat and diagnose pemphigus vulgaris  
PS Claim 1; Page 7-9; 9pp; Japanese.  
CC W07908 represents the human pemphigus vulgaris (PV) antigen  
CC extracellular region. The PV antigen is produced in patients with  
CC pemphigus vulgaris resulting in autoimmune disease. PV is a rare  
CC relapsing disease causing suppurative, intra-epidermal bullae  
CC (vesicles) of the skin and mucous membranes, which is fatal if  
CC untreated. The PV antigen was fused to a human IgG1 hinge region  
CC and the resulting fusion protein is useful to treat or diagnose

CC from the E2 initiation codon. The E2 and E3 genes have been  
CC cloned and can be used for the recombinant production of BCV  
CC polypeptides, using e.g. *Spodoptera frugiperda* Sf9 insect cells  
CC as host cells. Glycosylated and non-glycosylated recombinant  
CC E2 and E3 (see W31/07) are useful as components of vaccines  
CC directed toward preventing BCV infection, or reducing the severity  
CC of BCV infection, in bovine populations.  
SO Sequence 1363 AA:

Query Match	50.0%	Score 37	DB 1	Length 1363
Best Local Similarity	50.0%	Pred. No. 1.4e+02		
Matches	5	Conservative	4	Mismatches 1
				Indels 0
				Gaps 0

```
QY      5 ITRYISGVGI 14  
          : |||:|:|:  
Db      996 VQYRINGIGV 1005
```

## RESULT 15

AD Y04860 standard; Protein; 390 AA.  
AC Y04860;  
DT 06-JUL-1999 (first entry)  
DE Mycobacterium species protein sequence 18F.  
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;  
KW hybridisation; detection; vaccine; immunisation; infection.  
OS Mycobacterium sp.  
PN W0909186-A2.  
PD 25-FEB-1999.  
PE 14-AUG-1998; F01813.  
PR 11-SEP-1997; FR-011325.  
PR 14-AUG-1997; FR-010404.  
PI (INSP) INST PASTERE.  
PI Gicquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,  
PI Guignouet A;  
DR WPI: 99-181045/15.  
DR N-PSDB: X34112.  
PT Mycobacterial DNA vectors containing reporter constructs - for  
PT identifying coding or promoter sequences involved in  
PT infection-associated protein expression  
PS Claim 32: Fig 18F: 309pp; French.  
CC Sequences Y04742-Y05000 and Y07201-Y07204 represent secreted proteins  
CC from various Mycobacterium species microorganisms. The encoding  
CC nucleotide sequences can be used as primers and probes for methods  
CC for detecting and identifying mycobacteria, especially belonging to  
CC the M. tuberculosis complex. The encoded proteins can be used in  
CC vaccines for immunisation against a bacterial or viral infection.  
SQ Sequence 350 AA;

Query Match	50.0%;	Score 37;	DB 1;	Length 390;
Best Local Similarity	50.0%;	Pred. No. 35;		
Matches	6;	Conservative	4;	Mismatches 2; Indels 0; Gaps 0;

```
QY      4 KITYRISGVGID 15
          ::|||:|:||||
Db     10 QVTGRVFGIGID 21
```

Search completed: January 6, 2000, 15:40:34  
Job time: 1681 sec

PT Incorporated with its coding gene.  
 PS Claim 2; Fig 1f; 20pp; Japanese.  
 CC DNA encoding the protein can be inserted into an expression vector  
 CC for the prodn. of MG-1 polypeptide which elicits an antigen-antibody  
 CC reaction with anti-mycoplasma gallisepticum poultry sera. It may also  
 CC be ligated to other DNA to produce fusion proteins with an N-terminal  
 CC bacterial enzyme sequence.  
 CC See also R05081-2 and R06437-40.  
 SQ Sequence 309 AA;

Query Match 50.0%; Score 37; DB 1; Length 309;  
 Best Local Similarity 58.3%; Pred. No. 27;  
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 TORITRISGVG 13  
 DB 200 TOTVTNRFAGVG 211

RESULT 12  
 RI1061 R1061 standard; Protein: 1363 AA.  
 AC R1061:  
 DT 24-MAY-1991 (first entry)  
 DE Bovine Coronavirus E2 protein.  
 KW Bovine Coronavirus; BCV; E2; E3; vaccine.  
 OS Bovine coronavirus - Quebec isolate.  
 FH Key Location/Qualifiers  
 FT peptide 1..17  
 FT cleavage-site 766..766-sig\_peptide  
 FT /note="probable cleavage site"  
 FT domain 1306..1338  
 FT /label="transmembrane\_domain"  
 PN WO9102752-A.  
 PD 07-MAR-1991.  
 PF 10-AUG-1990; CA0252.  
 PR 22-AUG-1989; US-397689.  
 PA (VETE-) VETERINARY INFECTIO.  
 PI Parker MD, Cox GJ, Babluk LA;  
 DR WPI: 91-087247/12.  
 DR N-PSDB; Q10947.  
 PT Deoxyribonucleic acid encoding Bovine coronavirus protein E2 and  
 PT E3 - useful as vaccine component  
 PS Disclosure; Fig 3 (1-6); 70pp; English.  
 CC E2 protein is one of the four proteins (N, E1, E2 and E3) composing  
 CC BCV. The E2 and E3 polypeptides, or fragments of these, are used as  
 CC subunit antigens in vaccines for protection against or ameliorating  
 CC BCV, without risk of infection.  
 SQ Sequence 1363 AA;

Query Match 50.0%; Score 37; DB 1; Length 1363;  
 Best Local Similarity 50.0%; Pred. No. 14e+02;  
 Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 5 IYRISGVGI 14  
 DB 996 VOYRINGICV 1005

RESULT 13  
 WI13009 W13009 standard; protein: 560 AA.  
 AC W13009:  
 DT 21-NOV-1997 (first entry)  
 DE Segment of desmosomal cadherin, desmoglein Dsg2.  
 KW Desmosomal cadherin; desmoglein; Dsg2; cell; surface; epithelial;  
 KW carcinoma; desmosome; antibody; epitope; diagnosis; detection;  
 KW micrometastasis; separation; enrichment; targeted delivery;  
 KW metastatic.  
 OS Homo sapiens.

PN DE19531033-A1.  
 PD 27-FEB-1997.  
 PF 23-AUG-1995; 031033.  
 PR 23-AUG-1995; DE-031033.  
 PA (PROG-) PROGEN BIOTECHNIK GMBH.  
 PI Franke WW, Schaefer S;  
 DR WPI: 97-146518/14.  
 PT Antibody reactive with part of desmosomal cadherin - exposed on  
 PT surface of epithelial or carcinoma cells; not bound to desmosomes,  
 PT useful for diagnosis and treatment of carcinoma micrometastases  
 PS Claim 7; Page 5; 8pp; German.  
 CC The present sequence is a segment of the desmosomal cadherin (DC),  
 CC desmoglein Dsg2, which is exposed on the surface of epithelial or  
 CC carcinoma cells and not bound to desmosomes. An antibody (Ab)  
 CC directed against epitopes of the present sequence can be used to  
 CC diagnose, i.e. to detect carcinoma cells, especially  
 CC micrometastases, not bound to desmosomes, to separate, enrich or  
 CC detect living or fixed carcinoma cells by cell sorting methods and  
 CC as a therapeutic to deliver agents, e.g. other Ab or toxins, to  
 CC target cells. The Ab provides rapid and reliable detection of  
 CC metastatic carcinoma, and detects parts of DC that are not  
 CC accessible in desmosome bound cells, as in normal tissue or  
 CC carcinomas.  
 SQ Sequence 560 AA;

Query Match 50.0%; Score 37; DB 1; Length 560;  
 Best Local Similarity 63.6%; Pred. No. 52;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 4 KITRISGVGI 14  
 DB 35 KITRKTKGKI 45

RESULT 14  
 WI1706 W1706 standard; Protein: 1363 AA.  
 AC W1706:  
 DT 14-APR-1998 (first entry)  
 DE Bovine coronavirus E2 (S) protein.  
 KW BCV; E2 protein; peplomer protein; S spike; antigen; vaccine;  
 KW cattle.  
 OS Bovine coronavirus.  
 FH Key Location/Qualifiers  
 FT Peptide 1..17  
 FT /label="Sig\_peptide  
 FT Protein 18..1363  
 FT /label="Mat\_protein  
 FT Domain 1306..1338  
 FT /note="transmembrane domain"  
 PN US5672350-A.  
 PD 30-SEP-1997.  
 PF 22-AUG-1989; 397689.  
 PR 19-DEC-1991; US-811422.  
 PR 22-AUG-1989; US-397689.  
 PR 18-OCT-1991; US-779500.  
 PR 22-DEC-1993; US-171763.  
 PA (VETE-) VETERINARY INFECTIONOUS DISEASE.  
 PI Babluk LA, Cox GJ, Parker MD;  
 DR WPI: 97-468823/45.  
 DR N-PSDB; T89387.  
 PT Vaccines against bovine coronavirus - containing recombinant bovine  
 PT coronavirus polypeptide(s)  
 PS Claim 13; Fig 3; 52pp; English.  
 CC This polypeptide comprises the E2 protein, also designated  
 CC peplomer protein or S (spike), of bovine coronavirus (BCV). It  
 CC has a mol. wt. of 150 kDa exclusive of glycosylation and contains  
 CC 21 potential N-linked glycosylation sites. The amino acid  
 CC sequence was deduced from an clone E2 cDNA (see T89387). The  
 CC E2 gene in plasmid pT1822 (E. coli JM105) is deposited as ATCC  
 CC 68041. The BCV E3 gene (see T89388) is immediately 5' of the E2  
 CC gene on the viral genome and terminates 14 nucleotides upstream

PA (REGC ) UNIV CALIFORNIA.  
 PI Huang W, Nguyen TD, Polansky JR;  
 DR WPI; 98-542701/46.  
 N-PSDB; V33484.  
 PT New protein induced in trabecular meshwork cells by glucocorticoids  
 PS - useful in the diagnosis of glaucoma and related diseases  
 PS Claim 1, Fig 1A-C, 53pp, English.  
 CC This is the amino acid sequence of a new human 55 kDa protein,  
 CC designated trabecular meshwork induced glucocorticoid response\*  
 CC (TIGR\*) protein, that is highly induced by glucocorticoids in the  
 CC endothelial lining of the human trabecular meshwork (HTM). The  
 CC sequence was deduced from an isolated cDNA clone (see V33484).  
 CC Studies of the recombinant protein suggest (1) that the 55 kDa  
 CC protein exists both in cells and in the medium, (2) that it  
 CC undergoes oligomerisation, (3) phosphorylation, (4) glycosylation,  
 CC (5) that it is susceptible to metalloproteinase, (6) that it  
 CC exhibits high affinity binding to extracellular matrix and HTM  
 CC cells, (7) that it exhibits progressive inductions with time in  
 CC both cell and organ cultures, and (8) that it exhibits high  
 CC expression in the HTM of glaucomatous patients as compared to  
 CC normal patients. TIGR\* cDNA, the protein itself, molecules that  
 CC bind it, and nucleic acid molecules that encode it, provide  
 CC improved methods and reagents for diagnosing glaucoma and related  
 CC disorders, such as cardiovascular and immunological diseases that  
 CC affect expression of TIGR\*. A claimed method of diagnosing glaucoma  
 CC involves determining if the amount of TIGR\* present in the HTM  
 CC exceeds the amount found in an individual not predisposed to the  
 CC disease.  
 SQ Sequence 504 AA;

Query Match 52.7%; Score 39; DB 1; Length 504;  
 Best Local Similarity 57.1%; Pred. No. 20;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TOKITRISGVGID 15  
 ||:|:|:|:|:|  
 DB 281 TQETTRIDTVGTD 294

RESULT 9  
 ID W73500 standard; protein; 504 AA.  
 AC W73500;  
 DT 26-FEB-1999 (first entry)  
 DE Trabecular meshwork induced glucocorticoid response protein.  
 KW TIGR protein; trabecular meshwork induced glucocorticoid response;  
 KW secretory protein; antibody; glaucoma; diagnosis.  
 OS Homo sapiens.  
 PN US5849879-A.  
 PD 15-DEC-1998.  
 PF 14-MAY-1996; 645900.  
 PR 14-MAY-1996; US-645900.  
 PR 03-NOV-1994; US-336235.  
 PR 20-OCT-1995; US-546568.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Huang W, Nguyen TD, Polansky JR;  
 DR WPI; 99-069807/06.  
 PT Antibody to trabecular meshwork protein - useful for diagnosis of  
 PT glaucoma  
 PS Claim 1, Column 25-28; 22pp; English.  
 CC This sequence represents the human Trabecular meshwork induced  
 CC glucocorticoid response (TIGR) protein. The TIGR protein is a secretory  
 CC protein specifically bound by the antibody of the invention. The  
 CC antibody, especially in labeled form, can be used in the diagnosis of  
 CC glaucoma by detecting elevated levels of the protein in the trabecular  
 CC meshwork of the eye. Using the antibody, glaucoma is detected more  
 CC accurately.  
 SQ Sequence 504 AA;

Query Match 52.7%; Score 39; DB 1; Length 504;  
 Best Local Similarity 57.1%; Pred. No. 20;

Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TOKITRISGVGID 15  
 ||:|:|:|:|:|  
 DB 281 TQETTRIDTVGTD 294

RESULT 10  
 W89391  
 ID W89391 standard; Protein; 504 AA.  
 AC W89391;  
 DT 08-MAR-1999 (first entry)  
 DE Human trabecular meshwork induced glucocorticoid response protein.  
 KW Human; trabecular meshwork induced glucocorticoid response protein;  
 KW TIGR; glaucoma; primary open angle glaucoma; POG; pigmentary glaucoma;  
 KW low tension glaucoma; intraocular pressure; steroid; corticosteroid.  
 OS Homo sapiens.  
 PN US5854415-A.  
 PD 29-DEC-1998.  
 PF 25-JUN-1997; 882238.  
 PR 17-MAY-1996; US-649432.  
 PR 03-NOV-1994; US-336235.  
 PR 20-OCT-1995; US-546568.  
 PR 25-JUN-1997; US-882238.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Huang W, Nguyen TD, Polansky JR;  
 DR WPI; 99-095006/08.  
 DR N-PSDB; V81910.  
 PT New isolated glaucoma-associated nucleic acids - which encode  
 PT Trabecular Meshwork Induced Glucocorticoid Response protein, used to  
 PT develop products for diagnosing glaucoma-related diseases  
 PS Claim 1, Fig 1; 22pp; English.  
 CC The present sequence is a human secretory protein from clone II.2.  
 CC The secretory protein is designated TIGR (Trabecular Meshwork induced  
 CC Glucocorticoid Response) protein. The protein is highly induced by  
 CC glucocorticoids in the endothelial lining cells of the human trabecular  
 CC meshwork. The TIGR polynucleotides and proteins can be used as markers  
 CC for the diagnosis of glaucoma, primary open angle glaucoma (POAG),  
 CC pigmentary glaucoma, and low tension glaucoma and their related  
 CC diseases. They can also be used to diagnose or protect an individual's  
 CC sensitivity to elevated intraocular pressure upon administration of  
 CC steroids such as glucocorticoids or corticosteroids. These products can  
 CC also be used for diagnosing other diseases or conditions that affect the  
 CC expression or activity of the protein. The products can also be  
 CC formulated for administration to patients.  
 SQ Sequence 504 AA;

Query Match 52.7%; Score 39; DB 1; Length 504;  
 Best Local Similarity 57.1%; Pred. No. 20;  
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 TOKITRISGVGID 15  
 ||:|:|:|:|:|  
 DB 281 TQETTRIDTVGTD 294

RESULT 11  
 R06441  
 ID R06441 standard; protein; 309 AA.  
 AC R06441;  
 DT 08-OCT-1990 (first entry)  
 DE MG-6 antigen.  
 KW Mycoplasma gallisepticum; poultry; vaccine.  
 PN J02111795-A.  
 PD 24-APR-1990.  
 PF 02-JUN-1989; 136343.  
 PR 02-JUN-1989; JP-136343.  
 PA (JAPG) Nippon Zeon KK, (SHIO) Shionogi KK.  
 DR WPI; 90-169109/22  
 DR N-PSDB; 005653.  
 PT Diagnostic and vaccine for poultry mycoplasma serum - utilizes  
 PT antigen protein of the disease and recombinant vector



